TITLE

PARALLEL CHROMOSOMAL STACKING OF TRAITS IN BACTERIA

This application claims the benefit of U.S. Provisional Application No. 60/434,773 filed December 19, 2002.

FIELD OF THE INVENTION

This invention is in the field of microbiology. More specifically, this invention pertains to methods associated with *in vivo* chromosomal engineering.

BACKGROUND OF THE INVENTION

The availability of complete bacterial genome sequences and the elucidation of metabolic pathways have resulted in the use of such knowledge to engineer microorganisms for the production of compounds of industrial interest. Microbial production of industrial compounds requires the ability to efficiently engineer changes to the genomes of the organisms. Engineering changes such as adding, removing, or modifying genetic elements have often proven to be challenging and time consuming exercises. One such modification is genetically engineering modulations to the expression of relevant genes in a metabolic pathway.

There are a variety of ways to modulate gene expression. Microbial metabolic engineering generally involves the use of multi-copy vectors to express a gene of interest under the control of a strong or conditional promoter. This method of metabolic engineering for industrial use has several drawbacks. It is sometimes difficult to maintain the vector due to segregational instability. Deleterious effects on cell viability and growth are often observed due to the vector burden. It is also difficult to control the optimal expression level of desired genes on a vector. To avoid the undesirable effects of using a multi-copy vector, a general approach using homologous recombination via a single insertion of bacteriophage λ, transposons, or other suitable vectors containing the gene of interest has been used. However, this method also has drawbacks such as the need for multiple cloning steps in order to get the gene of interest into a suitable vector prior to recombination. Another drawback is the instability associated with the inserted genes, which can be lost due to excision. Lastly, these methods have a limitation associated with multiple insertions and the inability to control the location of the insertion site on a chromosome.

Although previous methods have been developed for making multiple DNA modifications in the chromosome, these have used transposons that are randomly integrated and require multiple cloning steps to insert genes of interest (Perdelchuk, M. Y., and Bennett, G. N. 1997. *Gene*. 187:231-238), or vectors that also require multiple cloning steps (PCT WO01/18222) and have not been applicable to all types of chromosomal modifications including insertions of whole genes or promoter sequences, deletions, and integrated transposons. Further, these methods have utilized a systematic approach to making multiple alterations at undefined loci as opposed to a combinatorial approach to making directed modifications on the chromosome.

The problem to be solved, therefore, is to define methods and materials to easily combine chromosomal modifications, created by any number of methods for chromosomal engineering, in one strain in a fashion that facilitates reaching optimum levels of product formation in bacteria, such as *E. coli.* The present invention has solved this problem by providing a method using P1 transduction and site-specific recombinase mediated marker excision to combine, in a linear, step-wise, and parallel combinatorial fashion chromosomal alterations. The present method allows for easy and efficient *in vivo* chromosomal engineering associated with biosynthetic pathway optimization.

SUMMARY OF THE INVENTION

The present method is a genetic tool useful for redesigning biosynthetic pathways, optimizing metabolic flux, and creating novel pathways by targeted *in vivo* chromosomal engineering. The method utilizes a homologous recombination system to introduce an integration cassette into a chromosome of a recombination proficient host cell and subsequently utilizes a phage transducing system to transfer the multiple integration cassettes into a single host cell in parallel combinatorial fashion. The "integration cassette" used to engineer the chromosomal modification includes a promoter and/or gene, and a selection marker flanked by site-specific recombinase sequences. After selection of the optimized transductants, a helper plasmid carrying a site-specific recombinase is introduced into the cells to excise the selectable markers bounded by site-specific recombinase sites. Repetition of the method facilitates combinatorial (multiple gene) trait stacking, necessary for biosynthetic pathway optimization.

The method can be used to engineer a variety of genetic elements, in addition to promoters, in the custom design of biosynthetic pathways. The approach is suitable for constructing industrially useful microbial strains, rather than just high expression of a specific single gene. In terms of metabolic balance, productivity, control, stability, and optimal expression of the genes of a particular pathway, the approach has many advantages and benefits when compared to metabolic engineering based on just a recombinant vector approach. The present method is illustrated using *E. coli* by example, but the method should prove to be useful in other bacterial strains as well.

The present method enables quick chromosomal trait stacking for optimal production of the desired genetic end product. A method that facilitates multiple chromosomal modifications is essential when engineering biosynthetic pathways for industrial purposes. The utility of the present method in engineering bacterial biosynthetic pathways is exemplified by altering isoprenoid and carotenoid biosynthesis. The promoters of the key genes encoding rate-limiting enzymes involved in the isoprenoid pathway (Figure 1) were engineered via the novel method. The genetic modifications accomplished by the present invention resulted in increased β-carotene production.

Accordingly the invention provides a method for the optimization of the production of a genetic end product comprising:

a) providing a multiplicity of integration cassettes, each cassette comprising:

(i) a nucleic acid integration fragment;

(ii) a selectable marker bounded by specific recombinase sites responsive to a recombinase;

(iii) homology arms having homology to different portions of a donor cell chromosome;

b) transforming at least one donor cell with the integration cassettes of (a) for its chromosomal integration;

c) infecting the transformed donor cell of (b) with a phage wherein the phage replicates and the donor cell is lysed;

d) isolating phage released by the lysis of the donor cell of (c);

e) mixing isolated phage released by the lysis the of donor cells of (c) carrying different integration cassettes of (a);

f) infecting a recipient cell with the mixture of the isolated phage of (e) wherein the integration cassettes each integrate

into the recipient cell chromosome at the point of homology to the homology arms to generate a transduced recipient cell;

g) selecting transduced recipient cells on the basis of the selectable marker;

h) screening the recipient cell of (g) for the highest level of the genetic end product to identify a first overproducing strain;

i) activating a recombinase in the first overproducing strain of (h) which excises the selectable marker from the chromosomally integrated integration cassette;

j) infecting the first over producing strain of (i) with the mixture of the isolated phage of (e) wherein the integration cassettes each integrate into the recipient cell chromosome at the point of homology on the homology arms;

k) screening the infected first overproducing strain of (j) for the highest level of the genetic end product to identify a second overproducing strain; and

l) comparing the levels of genetic end product produced by the first and second over producing strains whereby the production of the genetic end product is optimized.

In another embodiment the invention provides a method for the optimization of the production of a genetic end product comprising:

a) providing a multiplicity of integration cassettes, each cassette comprising:

(i) a promoter;

(ii) a selectable marker bounded by specific recombinase sites responsive to a recombinase;

(iii) regions of homology to different portions of a P1 donor cell chromosome;

b) transforming at least one donor cell with the integration cassette of (a) for its chromosomal integration;

c) infecting the transformed donor cell of (b) with a P1 phage wherein the phage replicates and the donor cell is lysed;

d) isolating phage released by the lysis of the donor cell of (c);

e) mixing equal number of isolating phage released by the lysis of a set of donor cells of (c) carrying different integration cassettes of (a);

f) infecting a recipient cell with the mixture of the isolated phage of (e) wherein the integration cassettes each integrate into the recipient cell chromosome at the point of homology to the homology arms;

g) selecting transduced recipient cells on the basis of the selectable marker;

h) screening the recipient cell of (f) for the highest level of the genetic end product to identify a first overproducing strain;

i) activating a recombinase in the first over producing strain of (h) which excises the selectable marker from the chromosomally integrated integration cassette;

j) infecting the first over producing strain of (i) with the mixture of the isolated phage of (e) wherein the integration cassettes each integrate into the recipient cell chromosome at the point of homology on the homology arms;

k) screening the first over producing strain of (j) for the highest level of the genetic end product to identify a second overproducing strain; and

l) comparing the levels of genetic end product produced by the first and second over producing strains whereby the production of the genetic end product is optimized.

BRIEF DESCRIPTION OF THE DRAWINGS SEQUENCE DESCRIPTIONS, AND BIOLOGICAL DEPOSITS

Figure 1 illustrates the isoprenoid/carotenoid biosynthetic pathway.

Figure 2 illustrates method of the invention for *in vivo* chromosomal engineering of stacking traits in a parallel combinatorial fashion.

Figure 3 illustrates the method of the invention for chromosomally integrating linear DNA using one or two PCR fragments.

Figure 4 illustrates the features of plasmid pSUH5.

Figure 5 illustrates the features of plasmid pKD46.

Figure 6 illustrates the features of plasmid pPCB15.

Figure 7 illustrates the scheme for increasing β-carotene levels in *E. coli* via the method of the invention.

Figure 8 illustrates elimination of the kanamycin resistance marker from the chromosome and an agarose gel verifying chromosomal integrations.

Figure 9 illustrates increased levels of β-carotene yielded by *E. coli* strains engineered via the present method.

The following biological deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the purposes of Patent Procedure:

Depositor Identification Reference	Int'l. Depository Designation	Date of Deposit
Plasmid pCP20	ATCC# PTA-4455	June 13, 2002

As used herein, "ATCC" refers to the American Type Culture Collection International Depository Authority located at ATCC, 10801 University Blvd., Manassas, VA 20110-2209, USA. The "International Depository Designation" is the accession number to the culture on deposit with ATCC.

The listed deposits will be maintained in the indicated international depository for at least thirty (30) years and will be made available to the public upon the grant of a patent disclosing it. The availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

The following sequences comply with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NOs:1-12 are nucleic acid and amino acid sequences encoding genes from the *Pantoea stewartii* carotenoid gene cluster.

SEQ ID NOs:13-32 are oligonucleotide primers used to create integrative fragments.

SEQ ID NOs:33-34 are oligonucleotide primers designed to amplify the carotenoid gene cluster from *Pantoea stewartii*.

SEQ ID NOs:35-39 are oligonucleotide primers used for screening for the presence of chromosomally integrated fragments.

SEQ ID NO:40 is the nucleotide sequence for plasmid pPCB15

SEQ ID NO:41 is the nucleotide sequence for plasmid pKD46.

SEQ ID NO:42 is the nucleotide sequence for plasmid pSUH5.

SEQ ID NO:43 is the nucleotide sequence for the bacteriophage *T5* promoter "P_{T5}".

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions, which form a part of this application.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides a process to efficiently combine multiple chromosomal modifications into a microorganism in order to optimize the production of a desired genetic end product. The process begins with chromosomally engineering alterations to individual genes known to be associated with a biosynthetic pathway. This first step is exemplified by chromosomally engineering changes to genes in the isoprenoid biosynthetic pathway by replacing the native gene promoters with the strong phage *T5* promoter (P_{T5}) using λ-Red mediated homologous recombination.

The λ-Red recombinase system facilitates efficient homologous recombination using linear DNA fragments having short regions (10-100 base pairs) of homology ("homology arms") to the targeted integration site. The integration fragment, termed "integration cassette", used to engineer the chromosomal modification includes a promoter and/or gene, and a selection marker flanked by site-specific recombinase sequences. Transformants are identified by incorporation of the selectable marker.

In a preferred embodiment, the bacteriophage P1 transduction system is used. Conventional P1 transduction can move only one genetic trait (i.e. gene) at a time from one host to another host. The present invention provides for a method moving multiple genetic traits into an *E. coli* host in a parallel combinatorial fashion using pooled mixtures of bacteriophage P1 in combination with a site-specific recombinase for removal of selection markers (Figure 2).

Referring to Figure 2, the various transformants (donor cells) are infected with phage P1. The resultant P1 lysates made from the various individual transformants are mixed. The integration fragments are randomly packed into phage particles which are subsequently used to infect a recipient cell, usually of the same species as that of the donor cell. Transduction and homologous recombination occurs, creating colonies containing various chromosomal integrations of the previously modified promoter and/or gene. The transduced recipient cells are screened for antibiotic resistance and assayed for increased production of the desired

genetic end product. After selection of the optimized transductants, the antibiotic resistance marker is removed by a site-specific recombinase. The selected transductants can be used again as a recipient cell in additional rounds of P1 transduction in order to engineer multiple chromosomal modifications, optimizing the production of the desired genetic end product.

In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

The term "genetic end product" means the substance, chemical or material that is produced as the result of the activity of a gene product. Typically a gene product is an enzyme and a genetic end product is the product of that enzymatic activity on a specific substrate. A genetic end product may be the result of a single enzyme activity or the result of a number of linked activities (i.e. an enzyme pathway).

The terms "stacking", "stacking traits", "parallel chromosomal stacking", and "trait stacking" are used interchangeably and refer to the repeated process of stacking multiple genetic traits into one *E. coli* host in parallel using bacteriophage P1 mixtures in combination with the site-specific recombinase system for removal of the selection markers (Figure 2).

The term "parallel combinatorial fashion" refers to the P1 transduction with the P1 lysate mixture made from various donor cells containing various genetic traits so that multiple genetic traits can be moved to the recipient cell in parallel.

The term "integration cassette" refers to a linear nucleic acid construct useful for the transformation of a recombination proficient bacterial host. Integration cassettes of the invention may include a variety of genetic elements such as selectable markers, expressible DNA fragments, and recombination regions having homology to regions on a bacterial chromosome or on other integration cassettes. Within the context of the present invention typically two integration cassettes are used for integration each with a single region of homology or "homology arm" to a portion if a bacterial chromosomal region.

The term "expressible DNA fragment" means any DNA that influences phenotypic changes in the host cell. An "expressible DNA fragment" may include for example, DNA comprising regulatory elements, isolated promoters, open reading frames, genes, or combinations thereof.

The terms "homology arm" and "recombination region" are used interchangeably and refer to a nucleotide sequence that enables homologous recombination between two nucleic acids having substantially the same nucleotide sequence in a particular region of two different nucleic acids. The preferred size range of the nucleotide sequence of the homology arm is from about 10 to about 100 nucleotides, where about 50 bp is preferred. Typically the level of base identity (defined herein as a one to one correspondence between the bases of each region) between the homology arm and the region of homology on the chromosome is at least about 70% where at least about 80% is preferred and where at least about 90% identity is most preferred.

The term "site-specific recombinase" is used in the present invention to describe a system comprised of one or more enzymes which recognize specific nucleotide sequences (recombination target sites) and which catalyze recombination between the recombination target sites. Site-specific recombination provides a method to rearrange, delete, or introduce exogenous DNA. Examples of site-specific recombinases and their associated recombination target sites include, but are not limited to *Cre*-lox, FLP/*FRT*, R/RS, Gin/gix, Xer/*dif,* Int/*att*, a pSR1 system, a cer system, and a fim system. The present invention illustrates the use of a site-specific recombinase to remove selectable markers. Antibiotic resistance markers, flanked on both sides by *FRT* recombination target sites, are removed by expression of the FLP site-specific recombinase.

The term "donor cell" refers to a bacterial strain susceptible to infection by a bacteriophage or virus, and which serves as a source for the nucleic acid fragments packaged into the transducing particles. Typically the genetic make up of the donor cell is similar or identical to the "recipient cell" which serves to receive lysate containing transducing phage or virus produced by the donor cell. As used herein, "P1 donor cell" is a bacterial strain susceptible to infection by a P1 bacteriophage.

The term "recipient cell" refers to a bacterial strain susceptible to infection by a bacteriophage or virus and which serves to receive lysate containing transducing phage or virus produced by the donor cell. A "P1 recipient cell" is a bacterial strain susceptible to infection by a P1 bacteriophage.

The term "selectable marker" means a gene encoding a gene product that, when present, enables one to identify and preferentially propagate a particular cell type.

The term "recombination proficient bacterial host" is used to describe a bacterial host which is capable of homologous recombination at rates useful for genetic engineering.

The term "homology" as applied to recombination regions and corresponding regions on a bacterial chromosome means nucleotide sequences sharing identical or nearly identical sequences. Complementary sequences between regions on the bacterial chromosome and recombination regions can associate and undergo homologous recombination in the presence of a recombinase system (i.e. λ-Red recombinase).

The terms "λ-Red recombination system", and "λ-Red system" are used interchangeably to describe a group of enzymes residing on a set of plasmids encoded by the bacteriophage λ genes *exo*, *bet*, and *gam*. The enzymes encoded by the three genes work together to increase the rate of homologous recombination in *E. coli*, an organism generally considered to have a relatively low rate of homologous recombination; especially when using linear integration cassettes. The λ-Red system facilitates the ability to use short regions of homology (10-50 bp) flanking linear dsDNA fragments for homologous recombination (Datsenko and Wanner, *PNAS*, 97:6640-6645 (2000)).

As used herein, the term "upstream" (when used in reference to a region of DNA) means the 5' side of a particular gene or sequence of nucleotides.

As used herein, the term "downstream" (when used in reference to a region of DNA) means the 3' side of a particular gene or sequence of nucleotides.

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a

native gene, comprising regulatory and coding sequences that are not found together in nature. The present invention illustrates the ability to genetically engineer replacement of a native gene's promoter with the phage *T5* ("P_{T5}") strong promoter. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Operon", in bacterial DNA, is a cluster of contiguous genes transcribed from one promoter that gives rise to a polycistronic mRNA.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed only in response to different environmental or physiological conditions are commonly referred to as "inducible promoters". Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters".

Promoters can also be categorized by the relative strength of their observed expression pattern (i.e. "weak", "moderate", "strong"). It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic", "recombinant" or "transformed" organisms.

The terms "transduction", "generalized transduction" and "P1 transduction" are used interchangeably and refer to a phenomenon in which bacterial DNA is transferred from one bacterial cell (the donor) to another (the recipient) by a phage particle containing bacterial DNA.

The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence

for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The term "isoprenoid" or "terpenoid" refers to the compounds and any molecules derived from the isoprenoid pathway including 10 carbon terpenoids and their derivatives, such as carotenoids and xanthophylls.

The term "Dxs" refers to the enzyme D-1-deoxyxylulose 5-phosphate encoded by the *dxs* gene that catalyzes the condensation of pyruvate and D-glyceraldehyde 3-phosphate to D-1-deoxyxylulose 5-phosphate (DOXP).

The terms "Dxr" or "IspC" refer to the enzyme DOXP reductoisomerase encoded by the *dxr* or *ispC* gene that catalyzes the simultaneous reduction and isomerization of DOXP to 2-C-methyl-D-erythritol-4-phosphate. The names of the gene, *dxr* or *ispC,* are used interchangeably in this application. The names of gene product, dxr or IspC are used interchangeably in this application.

The term "YgbP" or "IspD" and refers to the enzyme encoded by the *ygbB* or *ispD* gene that catalyzes the CTP-dependent cytidylation of 2-C-methyl-D-erythritol-4-phosphate to 4-diphosphocytidyl-2C-methyl-D-erythritol. The names of the gene, *ygbP* or *ispD,* are used interchangeably in this application. The names of gene product, YgbP or IspD are used interchangeably in this application.

The term "YchB" or "IspE" and refers to the enzyme encoded by the ychB or ispE gene that catalyzes the ATP-dependent phosphorylation of 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate. The names of the gene, *ychB* or *ispE,* are used interchangeably in this application. The names of gene product, YchB or IspE are used interchangeably in this application.

The term "YgbB" or "IspF" refers to the enzyme encoded by the *ybgB* or *ispF* gene that catalyzes the cyclization with loss of CMP of 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate to 2C-methyl-D-erythritol-2,4-cyclodiphosphate. The names of the gene, *ygbB* or *ispF,* are used interchangeably in this

application. The names of gene product, YgbB or IspF are used interchangeably in this application.

The term "GcpE" or "IspG" refers to the enzyme encoded by the *gcpE* or *ispG* gene that is involved in conversion of 2C-methyl-D-erythritol-2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate. The names of the gene, *gcpE* or *ispG,* are used interchangeably in this application. The names of gene product, GcpE or IspG are used interchangeably in this application.

The term "LytB" or "IspH" refers to the enzyme encoded by the *lytB* or *ispH* gene and is involved in conversion of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). The names of the gene, *lytB* or *ispH,* are used interchangeably in this application. The names of gene product, LytB or IspH are used interchangeably in this application.

The term "idi" refers to the enzyme isopentenyl diphosphate isomerase encoded by the *idi* gene that converts isopentenyl diphosphate to dimethylallyl diphosphate.

The term "ispA" refers to the enzyme farnesyl pyrophosphate (FPP) synthase encoded by the *ispA* gene.

The term "ispB" refers to the enzyme octaprenyl diphosphate synthase, which supplies the precursor of the side chain of the isoprenoid quinones encoded by the *ispB* gene (Figure 1).

The term "pPCB15" refers to the plasmid (Figure 6; SEQ ID NO:40) containing β-carotene synthesis genes *Pantoea crtEXYIB*, using as a reporter plasmid for monitoring β-carotene production in *E. coli* that is genetically engineered via the invented method.

The term "pKD46" refers to the helper plasmid expressing the λ-Red recombinase system comprising three essential genes, *exo*, *bet*, and *gam* (Figure 5; Datsenko and Wanner, *supra*; SEQ ID NO:41).

The term "pCP20" is a helper plasmid encoding the FLP site-specific recombinase (ATCC PTA-4455; Cherepanov and Wackernagel, *Gene*, 158:9-14 (1995); Datsenko and Wanner, *supra*).

The term "pSUH5" refers to the plasmid (Figure 4; SEQ ID NO:42) that was constructed in this invention by cloning a phage T5 promoter (P_{T5}) region into the *Nde*I restriction endonuclease site of pKD4 (Datsenko and Wanner, *supra*). It was used as a template plasmid for PCR amplification of a fused kanamycin selectable marker/phage *T5* promoter linear DNA nucleotide.

The terms "P_{T5} promoter", "phage *T5* promoter", and "P_{T5}" refer to the nucleotide sequence that comprises the –10 and –35 consensus sequences, lactose operator (*lacO*), and ribosomal binding site (*rbs*) from phage *T5* (SEQ ID NO:43).

The term "helper plasmid" refers to either pKD46 encoding λ-Red recombinase or pCP20 (ATCC PTA-4455) encoding FLP site-specific recombinase (Cherepanov and Wackernagel, *supra;* Datsenko and Wanner, *supra*).

The term "*E. coli*" refers to *Escherichia coli* strain K-12 derivatives, such as MG1655 (ATCC 47076) and MC1061 (ATCC 53338).

The term "*Pantoea stewartii* subsp. *stewartii*" is abbreviated as "*Pantoea stewartii*" and is used interchangeably with *Erwinia stewartii* (Mergaert et al., *Int J. Syst. Bacteriol.*, 43:162-173 (1993)).

The term "*Pantoea ananatas*" is used interchangeably with *Erwinia uredovora* (Mergaert et al., *supra*).

The term "*Pantoea crtEXYIB* cluster" refers to a gene cluster containing carotenoid synthesis genes *crtEXYIB* amplified from *Pantoea stewartii* ATCC 8199. The gene cluster contains the genes *crtE*, *crtX*, *crtY*, *crtI*, and *crtB*. The cluster also contains a *crtZ* gene organized in opposite direction and adjacent to *crtB* gene.

The term "CrtE" refers to geranylgeranyl pyrophosphate (GGPP) synthase enzyme encoded by *crtE* gene which converts trans-trans-farnesyl diphosphate + isopentenyl diphosphate to pyrophosphate + geranylgeranyl diphosphate.

The term "CrtY" refers to lycopene cyclase enzyme encoded by *crtY* gene which converts lycopene to β-carotene.

The term "CrtI" refers to phytoene dehydrogenase enzyme encoded by *crtI* gene which converts phytoene into lycopene via the intermediaries of phytofluene, zeta-carotene and neurosporene by the introduction of 4 double bonds

The term "CrtB" refers to phytoene synthase enzyme encoded by *crtB* gene which catalyzes reaction from prephytoene diphosphate (geranylgeranyl pyrophosphate) to phytoene.

The term "CrtX" refers to zeaxanthin glucosyl transferase enzyme encoded by *crtX* gene which converts zeaxanthin to zeaxanthin-β-diglucoside.

The term "CrtZ" refers to the β-carotene hydroxylase enzyme encoded by *crtZ* gene which catalyses hydroxylation reaction from β-carotene to zeaxanthin.

The term "isoprenoid biosynthetic pathway" refers to those genes comprising members of the upper and/or lower isoprenoid pathways of the present invention as illustrated in Figure 1. In the present invention, the terms "upper isoprenoid pathway" and "upper pathway" will be use interchangeably and will refer the enzymes involved in converting pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP). These enzymes include, but are not limited to Dxs, Dxr (IspC), YgpP (IspD), YchB (IspE), YgbB (IspF), GcpE (IspG), LytB (IspH), Idi, IspA, and optionally IspB. In the present invention, the terms "lower isoprenoid pathway", "carotenoid biosynthetic pathway", and "lower pathway" will be used interchangeably and refer to those enzymes which convert FPP to carotenes, especially β-carotene (Figure 1). The enzymes in this pathway include, but are not limited to CrtE, CrtY, CrtI, CrtB, CrtX, and CrtZ. In the present invention, the "lower pathway" genes are expressed on a reporter plasmid, pPCB15.

The terms "carotenoid biosynthetic enzyme" is an inclusive term referring to any and all of the enzymes encoded by the *Pantoea crtEXYIB* cluster. The enzymes include CrtE, CrtY, CrtI, CrtB, and CrtX.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA), and the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-120. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY. Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized.

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) (hereinafter "Maniatis"); and by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory Cold Press Spring Harbor, NY (1984); and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience (1987).

Integration Cassettes

As used in the present invention, "integration cassettes" are the linear double-stranded DNA fragments chromosomally integrated by homologous recombination via the use of two PCR-generated fragments or one PCR-generated fragment as seen in Figure 3. The integration cassette comprises a nucleic acid integration fragment that contains a promoter and/or expressible DNA fragment and a selectable marker bounded by specific recombinase sites responsive to a site-specific recombinase and homology arms having homology to different portions of a donor cell chromosome. Typically, the integration cassette will have the general structure: 5'-RR1-RS-SM-RS-Y-RR2-3' wherein

(i) RR1 is a first homology arm of about 10 to 100 bases;

(ii) RS is a recombination site responsive to a site-specific recombinase;

(iii) SM is a DNA fragment encoding a selectable marker;

(iv) Y is a first expressible DNA fragment; and

(v) RR2 is a second homology arm.

Expressible DNA fragments of the invention are those that will be useful in the genetic engineering of pathways. For example, it may be useful to engineer a strong promoter in place of a native promoter in certain pathways. Virtually any promoter is suitable for the present invention including, but not limited to *lac*, *ara*, *tet*, *trp*, λP_L, λP_R, *T7*, *tac*, P_{T5}, and *trc* (useful for expression in *Escherichia coli*) as well as the *amy*, *apr*, *npr* promoters and various phage promoters useful for expression in *Bacillus*, for example.

Typically the invention makes use of at least two integration cassettes, each having a single region of homology or homology arm to a portion of the chromosome. The use of two cassettes in this conformation is illustrated in Figure 3 and will be referred to herein as the "two fragment

method". The two fragment method provides high rates of integration and is useful for the elimination of additional cloning steps.

Alternatively, different coding regions may be introduced downstream of existing native promoters. In this manner new coding regions encoding members of a biosynthetic pathway may be introduced that add, remove, decrease, or enhance the desired activity of the targeted biosynthetic pathway. The biosynthetic pathway can either foreign or endogenous to the host cell. Preferably, one or more members of the biosynthetic pathway already exist in the host cell. These coding regions may be genes which retain their native promoters or may be chimeric genes operably linked to an inducible or constitutive strong promoter for increased expression of the genes in the targeted biosynthetic pathway. Preferred in the present invention are the genes of the isoprenoid and/or carotenoid biosynthetic pathway, which include *dxs*, *dxr*, *ygbP*, *ychB*, *ygbB*, *idi*, *ispA*, *lytB*, *gcpE*, *ispB*, *gps*, *crtE*, *crtY*, *crtI*, *crtB*, *crtX*, and *crtZ*, as defined above and illustrated in Figure 1. In some situations the expressible DNA fragment may be in antisense orientation where it is desired to down-regulate certain elements of the pathway.

In the present invention it is preferred if the expressible DNA fragment is a promoter or a coding region useful for modulation of a biosynthetic pathway. Exemplified in the invention is the phage *T5* promoter used for the modulation of the isoprenoid biosynthetic pathway in a recombination proficient *E. coli* host.

Generally, the preferred length of the homology arms is about 10 to about 100 base pairs in length, where about 50 bp is preferred. Given the relatively short lengths of the homology arms used in the present invention for homologous recombination, one would expect that the level of acceptable mismatched sequences should be kept to an absolute minimum for efficient recombination, preferably using sequences which are identical to those targeted for homologous recombination. From 20 to 40 base pairs of homology, the efficiency of homologous recombination increases by four orders of magnitude (Yu et al., *PNAS*, 97:5978-5983 (2000)). Therefore, multiple mismatching within homology arms may decrease the efficiency of homologous recombination; however, one skilled in the art can easily ascertain the acceptable level of mismatching.

The present invention makes use of a selectable marker on one of the two integration cassettes ("two fragment method"). Numerous selectable markers are known to those skilled in the art. The selectable

marker is selected from the group consisting of antibiotic resistance markers, enzymatic markers (whereby the expressed marker catalyzes a chemical reaction creating a measurable difference in phenotypic appearance, for example, β-galactosidase), and amino acid biosynthesis enzymes which enable a normally auxotrophic bacteria to grow without the exogenously supplied amino acid. Examples of antibiotic resistance markers include ampicillin (amp^r), kanamycin (kan^r), and tetracycline (tet^r) resistance, to name a few. As used herein, the selectable markers are flanked by site-specific recombinase recognition sequences. After selection and construct verification, a site-specific recombinase is used to remove the selectable marker. The steps of the present method can then be repeated for additional *in vivo* chromosomal modifications. The integration cassette is bounded by site-specific recombinases for the eventual removal of the selectable marker. Site-specific recombinases, such as flippase (FLP) recombinase in the present invention, recognize specific recombination sequences (i.e. *FRT* sequences) and excise of the selectable marker. This aspect of the invention enables the repetitive use of the present method for multiple chromosomal modifications. The invention is not limited to the FLP-*FRT* recombinase system as several examples of site-specific recombinases and their associated specific recognition sequences are known in the art. Examples of other suitable site-specific recombinases and their corresponding recognition sequences include, but are not limited to *Cre*-lox, R/RS, Gin/gix, Xer/*dif,* Int/*att*, a pSR1 system, a cer system, and a fim system.

Recombination Proficient Host Cells

The present invention makes use of a recombination proficient host cell that is able to mediate efficient homologous recombination between the two integration cassettes and the host cell chromosome. Some organisms mediate homologous recombination very effectively (yeast for example) while others require genetic intervention. For example, *E. coli*, a host generally considered as one that does not undergo efficient transformation via homologous recombination naturally, may be altered to make it a recombination proficient host. Transformation with a helper plasmid containing the λ-Red recombinase system increases the rate of homologous recombination several orders of magnitude (Murphy et al., *Gene*, 246:321-330 (2000); Murphy, K., *J. Bacteriol.*, 180:2063-2071; Poteete and Fenton, *J. Bacteriol.*, 182:2336-2340 (2000); Poteete, A., *FEMS Microbiology Lett.*, 201:9-14 (2001); Datsenko and Wanner, *supra*;

Yu et al., *supra*; Chaveroche et al., *Nucleic Acids Research*, 28:e97:1-6 (2000); US 6,355,412; US 6,509,156; and US SN 60/434602). The λ-Red system can also be chromosomally integrated into the host. The λ-Red system contains three genes (*exo*, *bet*, and *gam*) which change the normally recombination deficient *E. coli* into a recombination proficient host.

Normally, *E. coli* efficiently degrade linear double-stranded (ds) DNA via its RecBCD endonuclease, resulting in transformation efficiencies not useful for chromosomal engineering. The *gam* gene encodes for a protein that binds to the *E. coli* RecBCD complex, inhibiting the undesirable endonuclease activity. The *exo* gene encodes for a λ-exonuclease that processively degrades the 5' end strand of double-stranded dsDNA and creates 3' single stranded overhangs. The protein encoded by *bet* complexes with the λ-exonuclease and binds to the single-stranded DNA overhangs and promotes renaturation of complementary strands and is capable of mediating exchange reactions. The λ-Red recombinase system enables the use of homologous recombination as a tool for *in vivo* chromosomal engineering in hosts normally considered difficult to transform by homologous recombination. The λ-Red system works in other bacteria as well (Poteete, A., *supra*, (2001)). The λ-Red system should be applicable to other hosts generally used for industrial production. These additional hosts include, but are not limited to *Agrobacterium, Erythrobacter, Chlorobium, Chromatium, Flavobacterium, Cytophaga, Rhodobacter, Rhodococcus, Streptomyces, Brevibacterium, Corynebacteria, Mycobacterium, Deinococcus, Paracoccus, Escherichia, Bacillus, Myxococcus, Salmonella, Yersinia, Erwinia, Pantoea, Pseudomonas, Sphingomonas, Methylomonas, Methylobacter, Methylococcus, Methylosinus, Methylomicrobium, Methylocystis, Alcaligenes, Synechocystis, Synechococcus, Anabaena, Thiobacillus, Methanobacterium, Klebsiella,* and *Myxococcus*.

<u>λ-Red Recombinase System</u>

The λ-Red recombinase system used in the present invention is contained on a helper plasmid (pKD46) and is comprised of three essential genes, *exo*, *bet*, and *gam* (Datsenko and Wanner, *supra)*. The *exo* gene encodes an λ-exonuclease, which processively degrades the 5' end strand of double-stranded (ds) DNA and creates 3' single-stranded overhangs. *Bet* encodes for a protein which complexes with the λ-exonuclease and binds to the single stranded DNA and promotes

renaturation of complementary strands and is capable of mediating exchange reactions. *Gam* encodes for a protein that binds to the *E.coli*'s RecBCD complex and blocks the complex's endonuclease activity.

The λ-Red system is used in the present invention because homologous recombination in *E.coli* occurs at a very low frequency and usually requires extensive regions of homology. The λ-Red system facilitates the ability to use short regions of homology (10-50 bp) flanking linear dsDNA fragments for homologous recombination. Additionally, the RecBCD complex normally expressed in *E.coli* prevents the use of linear dsDNA for transformation as the complex's exonuclease activity efficiently degrades linear dsDNA. Inhibition of the RecBCD complex's endonuclease activity by *gam* is essential for efficient homologous recombination using linear dsDNA fragments.

Combinatorial P1 Transduction System

Transduction is a phenomenon in which bacterial DNA is transferred from one bacterial cell (the donor) to another (the recipient) by a phage particle containing bacterial DNA. When a population of donor bacteria is infected with a phage, the events of the phage lytic cycle may be initiated. During lytic infection, the enzymes responsible for packaging viral DNA into the bacteriophage sometimes package host DNA. The resulting particle is called a transducing particle. Upon lysis of the cell, these particles are released along with the normal virions. The lysate contains a mixture of normal virions and transducing particles. When the lysate is used to infect a population of recipient cells, most of the cells become infected with normal virus. However, a small proportion of the population receives transducing particles that inject the DNA they received from the previous host bacterium. This DNA can now undergo genetic recombination with the DNA of the recipient host. Conventional use of P1 transduction can move only one genetic trait (i.e. gene) at a time from one host to another.

It will be appreciated that a number of host systems may be used for purposes of the present invention including, but not limited to those with known transducing phages such as *Agrobacterium, Erythrobacter, Chlorobium, Chromatium, Flavobacterium, Cytophaga, Rhodobacter, Rhodococcus, Streptomyces, Brevibacterium, Corynebacteria, Mycobacterium, Deinococcus, Paracoccus, Escherichia, Bacillus, Myxococcus, Salmonella, Yersinia, Erwinia, Pantoea, Pseudomonas, Sphingomonas, Methylomonas, Methylobacter, Methylococcus,*

Methylosinus, Methylomicrobium, Methylocystis, Alcaligenes, Synechocystis, Synechococcus, Anabaena, Thiobacillus, Methanobacterium, Klebsiella, and *Myxococcus*. Examples of phages suitable for use in the present invention include P1, P2, lambda, ϕ80, ϕ3538, T1, T4, P22, P22 derivatives, ES18, Felix "o", P1-CmCs, Ffm, PY20, Mx4, Mx8, PBS-1, PMB-1, and PBT-1.

The present method provides a system for moving multiple genetic traits into an *E. coli* host in a parallel combinatorial fashion using the bacteriophage P1 lysate in combination with the site-specific recombinase system for removal of selection markers (Figure 2). After transduction with the P1 lysate mixture made from various donor cells, the transduced recipient cells are screened for antibiotic resistance and assayed for increased production of the desired genetic end product (i.e. carotenoid production). After selection of the optimized transductants, the antibiotic resistance marker is removed by a site-specific recombinase. The selected transductants are used again as a recipient cell in additional rounds of P1 transduction in order to engineer multiple chromosomal modifications, optimizing production of the desired genetic end product. The present combinatorial P1 transduction method enables quick chromosomal trait stacking for optimal production of the desired genetic end product. The invention is very useful for continuous strain improvement toward a desired product.

Biosynthetic Pathway Optimization

The present method of combinatorial P1 transduction is applicable to the optimization of any biosynthetic pathway including isoprenoids, terpenoids, tetrapyrroles, polyketides, vitamins, amino acids, fatty acids, proteins, nucleic acids, carbohydrates, antimicrobial agents, anticancer agents and biological metabolites, to name a few.

The utility of the present invention is specifically illustrated by optimizing the isoprenoid and carotenoid biosynthetic pathways. Specifically, the method was used to identify the *ispB* gene by measuring its effects on the production of β-carotene. Carotenoid production (i.e. β-carotene) was enhanced by operably linking the phage *T5* promoter to the coding sequence of the gene.

It will be appreciated that another pathway amenable to engineering by the present method is the heme biosynthetic pathway. One skilled in the art can prepare a series of strains where the promoters for each of the genes involved in the synthesis of tetrapyrroles, such as

heme, are replaced by inducible promoters as described in this invention. Examples of the heme synthesis genes include for example: *hemA, hemL, hemB, hemC, hemD, hemE, hemF, hemG*, and *hemH.* The first step is chromosomally engineering changes to genes in the heme biosynthetic pathway by replacing the native gene promoters with a foreign promoter using λ-Red mediated homologous recombination. Next, combinatorial P1 transduction using P1 mixtures (Figure 2) can be used to combinatorially stack the fused foreign promoter-heme genes for selecting a strain producing increased amounts of tetrapyrrole. One skilled in the art can select an appropriate promoter for use in the present method to optimize production of the desired genetic end product. The transductants are analyzed for production of heme and analyzed by traditional analytical methods, such as mass spectroscopy, UV-VIS spectrometry, bioassays or enzymatic coupled assays. A site-specific recombination system, such as the FLP/*FRT* or Cre-*lox* recombinase system, can be utilized to remove the markers from the strain after each stacking cycle. The process can be repeated, selecting those transductants for optimal production of the desired tetrapyrrole products.

Modulation of Genes Involved in Carotenoid Production.

The enzymatic pathway involved in the biosynthesis of carotenoids can be conveniently viewed in two parts, the upper isoprenoid pathway that facilitates the conversion of pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP) and the lower carotenoid biosynthetic pathway, which provides for the synthesis of phytoene and all subsequently produced carotenoids (Figure 1). The upper pathway is ubiquitous in many microorganisms. In these cases it will only be necessary to introduce genes that comprise the lower pathway for the biosynthesis of carotenoids. The key division between the two pathways concerns the synthesis of farnesyl pyrophosphate. Where FPP is naturally present, only elements of the lower carotenoid pathway will be needed. However, it will be appreciated that for the lower pathway carotenoid genes to be effective in the production of carotenoids, it will be necessary for the host cell to have suitable levels of FPP within the cell. Where the host cell does not provide a suitable level of FPP synthesis, it will be necessary to introduce chromosomal modifications (promoters, genes, etc.) necessary for the production of FPP. These modifications can be introduced in the host by using the present method. Each of these pathways will be discussed below in detail.

The Upper Isoprenoid Pathway

Isoprenoid biosynthesis occurs through either of two pathways, generating the common C5 isoprene subunit, isopentenyl pyrophosphate (IPP). First, isopentenyl pyrophosphate (IPP) may be synthesized through the well-known acetate/mevalonate pathway. However, recent studies have demonstrated that the mevalonate-dependent pathway does not operate in all living organisms. An alternate mevalonate-independent pathway for IPP biosynthesis has been characterized in bacteria and in green algae and higher plants (Horbach et al., *FEMS Microbiol. Lett.*, 111:135-140 (1993); Rohmer et al, *Biochem.*, 295: 517-524 (1993); Schwender et al., *Biochem.*, 316: 73-80 (1996); Eisenreich et al., *Proc. Natl. Acad. Sci. USA,* 93: 6431-6436 (1996)).

Many steps in the mevalonate-independent isoprenoid biosynthetic pathway are known (Figure 1). For example, the initial steps of the alternate pathway leading to the production of IPP have been studied in *Mycobacterium tuberculosis* by Cole et al. (*Nature,* 393:537-544 (1998)). The first step of the pathway involves the condensation of two 3-carbon molecules (pyruvate and D-glyceraldehyde 3-phosphate) to yield a 5-carbon compound known as D-1-deoxyxylulose-5-phosphate. This reaction occurs by the DXS enzyme, encoded by the *dxs* gene. Next, the isomerization and reduction of D-1-deoxyxylulose-5-phosphate yields 2-C-methyl-D-erythritol-4-phosphate. One of the enzymes involved in the isomerization and reduction process is D-1-deoxyxylulose-5-phosphate reductoisomerase (DXR), encoded by the gene *dxr (ispC)*. 2-C-methyl-D-erythritol-4-phosphate is subsequently converted into 4-diphosphocytidyl-2C-methyl-D-erythritol in a CTP-dependent reaction by the enzyme encoded by the non-annotated gene *ygbP*. The *ygbP* gene was recently renamed as *ispD* as a part of the *isp* gene cluster (SwissProtein Accession #Q46893).

Next, the 2nd position hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol can be phosphorylated in an ATP-dependent reaction by the enzyme encoded by the *ychB* gene, recently renamed *ispE* (SwissProtein Accession #P24209). This product phosphorylates 4-diphosphocytidyl-2C-methyl-D-erythritol, resulting in 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate. The product of the *ygbB* gene converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate in a CTP-dependent manner.

This gene has also been recently renamed as *isp* (SwissProtein Accession #P36663).

The enzymes encoded by the *gcpE* (*ispG*) and *lytB* (*ispH*) genes (and perhaps others) are thought to participate in the reactions leading to formation of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). IPP may be isomerized to DMAPP via IPP isomerase, encoded by the *idi* gene. However, this enzyme is not essential for survival and may be absent in some bacteria using 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Recent evidence suggests that the MEP pathway branches before IPP and separately produces IPP and DMAPP via the *lytB* (*ispH*) gene product. A *lytB* (*ispH*) knockout mutation is lethal in *E. coli* except in media supplemented with both IPP and DMAPP.

The synthesis of FPP occurs via isomerization of IPP to dimethylallyl pyrophosphate (DMAPP). This reaction is followed by a sequence of two prenyltransferase reactions catalyzed by *ispA,* leading to the creation of geranyl pyrophosphate (GPP; a 10-carbon molecule) and farnesyl pyrophosphate (FPP; 15-carbon molecule).

Genes encoding elements of the upper pathway are known from a variety of plant, animal, and bacterial sources, as shown in Table 1.

Table 1

Sources of Genes Encoding the Upper Isoprene Pathway

Gene	***GenBank Accession Number and Source Organism***
dxs (D-1-deoxyxylulose 5-phosphate synthase)	AF035440, *Escherichia coli* Y18874, *Synechococcus* PCC6301 AB026631, *Streptomyces sp.* CL190 AB042821, *Streptomyces griseolosporeus* AF111814, *Plasmodium falciparum* AF143812, *Lycopersicon esculentum* AJ279019, *Narcissus pseudonarcissus* AJ291721, *Nicotiana tabacum*
***dxr* (*ispC*)** (1-deoxy-D-xylulose 5-phosphate reductoisomerase)	AB013300, *Escherichia coli* AB049187, *Streptomyces griseolosporeus* AF111813, *Plasmodium falciparum* AF116825, *Mentha* x piperita AF148852, *Arabidopsis thaliana* AF182287, *Artemisia annua* AF250235, *Catharanthus roseus* AF282879, *Pseudomonas aeruginosa*

Gene	***GenBank Accession Number and Source Organism***
	AJ242588, *Arabidopsis thaliana* AJ250714, *Zymomonas mobilis* strain ZM4 AJ292312, *Klebsiella pneumoniae*, AJ297566, *Zea mays*
***ygbP* (*ispD*)** (2-C- methyl-D-erythritol 4-phosphate cytidylyltransferase)	AB037876, *Arabidopsis thaliana* AF109075, *Clostridium difficile* AF230736, *Escherichia coli* AF230737, *Arabidopsis thaliana*
***ychB* (*ispE*)** (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase)	AF216300, *Escherichia coli* AF263101, *Lycopersicon esculentum* AF288615, *Arabidopsis thaliana*
***ygbB* (*ispF*)** (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase)	AB038256, *Escherichia coli mecs* gene AF230738, *Escherichia coli* AF250236, *Catharanthus roseus* (MECS) AF279661, *Plasmodium falciparum* AF321531, *Arabidopsis thaliana*
***gcpE* (*ispG*)** (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase)	O67496, *Aquifex aeolicus* P54482, *Bacillus subtilis* Q9pky3, *Chlamydia muridarum* Q9Z8H0, *Chlamydophila pneumoniae* O84060, *Chlamydia trachomatis* P27433, *Escherichia coli* P44667, *Haemophilus influenzae* Q9ZLL0, *Helicobacter pylori* J99 O33350, *Mycobacterium tuberculosis* S77159, *Synechocystis* sp. Q9WZZ3, *Thermotoga maritima* O83460, *Treponema pallidum* Q9JZ40, *Neisseria meningitidis* Q9PPM1, *Campylobacter jejuni* Q9RXC9, *Deinococcus radiodurans* AAG07190, *Pseudomonas aeruginosa* Q9KTX1, *Vibrio cholerae*
***lytB* (*ispH*)**	AF027189, *Acinetobacter sp.* BD413 AF098521, *Burkholderia pseudomallei* AF291696, *Streptococcus pneumoniae* AF323927, *Plasmodium falciparum* gene M87645, *Bacillus subtillis* U38915, *Synechocystis sp.* X89371, *C. jejunisp* O67496
lspA (FPP synthase)	AB003187, *Micrococcus luteus* AB016094, *Synechococcus elongatus*

Gene	***GenBank Accession Number and Source Organism***
	AB021747, *Oryza sativa FPPS1* gene for farnesyl diphosphate synthase AB028044, *Rhodobacter sphaeroides* AB028046, *Rhodobacter capsulatus* AB028047, *Rhodovulum sulfidophilum* AF112881 and AF136602, *Artemisia annua* AF384040, *Mentha* x piperita D00694, *Escherichia coli* D13293, *B. stearothermophilus* D85317, *Oryza sativa* X75789, *A. thaliana* Y12072, *G. arboreum* Z49786, *H. brasiliensis* U80605, *Arabidopsis thaliana* farnesyl diphosphate synthase precursor (*FPS1*) mRNA, complete cds X76026, *K. lactis FPS* gene for farnesyl diphosphate synthetase, *QCR8* gene for bc1 complex, subunit VIII X82542, P.argentatum mRNA for farnesyl diphosphate synthase (FPS1) X82543, *P. argentatum* mRNA for farnesyl diphosphate synthase (FPS2) BC010004, *Homo sapiens*, farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase), clone MGC 15352 IMAGE, 4132071, mRNA, complete cds AF234168, *Dictyostelium discoideum* farnesyl diphosphate synthase (Dfps) L46349, *Arabidopsis thaliana* farnesyl diphosphate synthase (*FPS2*) mRNA, complete cds L46350, *Arabidopsis thaliana* farnesyl diphosphate synthase (*FPS2*) gene, complete cds L46367, *Arabidopsis thaliana* farnesyl diphosphate synthase (FPS1) gene, alternative products, complete cds M89945, Rat farnesyl diphosphate synthase gene, exons 1-8 NM_002004, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA U36376, *Artemisia annua* farnesyl diphosphate synthase (*fps1*) mRNA, complete cds XM_001352, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA

Gene	GenBank Accession Number and Sourc Organism
	XM_034497, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA XM_034498, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA XM_034499, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA XM_0345002, *Homo sapiens* farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (*FDPS*), mRNA

The Lower Carotenoid Biosynthetic Pathway

The division between the upper isoprenoid pathway and the lower carotenoid pathway is somewhat subjective. Because FPP synthesis is common in both carotenogenic and non-carotenogenic bacteria, the first step in the lower carotenoid biosynthetic pathway is considered to begin with the prenyltransferase reaction converting farnesyl pyrophosphate (FPP) to geranylgeranyl pyrophosphate (GGPP). The gene *crtE*, encoding GGPP synthetase is responsible for this prenyltransferase reaction which adds IPP to FPP to produce the 20-carbon molecule GGPP. A condensation reaction of two molecules of GGPP occurs to form phytoene (PPPP), the first 40-carbon molecule of the lower carotenoid biosynthesis pathway. This enzymatic reaction is catalyzed by *crtB*, encoding phytoene synthase.

Lycopene, which imparts a "red"-colored spectra, is produced from phytoene through four sequential dehydrogenation reactions by the removal of eight atoms of hydrogen, catalyzed by the gene *crtI* (encoding phytoene desaturase). Intermediaries in this reaction are phytofluene, zeta-carotene, and neurosporene.

Lycopene cyclase (*crtY*) converts lycopene to β-carotene.

β-carotene is converted to zeaxanthin via a hydroxylation reaction resulting from the activity of β-carotene hydroxylase (encoded by the *crtZ* gene). β-cryptoxanthin is an intermediate in this reaction.

β-carotene is converted to canthaxanthin by β-carotene ketolase encoded by a *crtW* or *crtO* gene. Echinenone in an intermediate in this reaction. Canthaxanthin can then be converted to astaxanthin by β-carotene hydroxylase encoded by a *crtZ* or *crtR* gene. Adonbirubrin is an intermediate in this reaction.

Zeaxanthin can be converted to zeaxanthin-β-diglucoside. This reaction is catalyzed by zeaxanthin glucosyl transferase (*crtX*).

Zeaxanthin can be converted to astaxanthin by β–carotene ketolase encoded by ketolase encoded by *crtW*, *crtO* or *bkt*. The BKT/CrtW enzymes synthesized canthaxanthin via echinenone from β-carotene and 4-ketozeaxanthin. Adonixanthin is an intermediate in this reaction.

Spheroidene can be converted to spheroidenone by spheroidene monooxygenase encoded by *crtA*.

Neurosporene can be converted spheroidene and lycopene can be converted to spirilloxanthin by the sequential actions of hydroxyneurosporene synthase, methoxyneurosporene desaturase and hydroxyneurosporene-O-methyltransferase encoded by the *crtC*, *crtD* and *crtF* genes, respectively.

β-carotene can be converted to isorenieratene by β-carotene desaturase encoded by *crtU* .

Genes encoding elements of the lower carotenoid biosynthetic pathway are known from a variety of plant, animal, and bacterial sources, as shown in Table 2.

Table 2

Sources of Genes Encoding the Lower Carotenoid Biosynthetic Pathway

Gene	***GenBank Accession Number and Source Organism***
crtE *(GGPP Synthase)*	AB000835, *Arabidopsis thaliana* AB016043 and AB019036, *Homo sapiens* AB016044, *Mus musculus* AB027705 and AB027706, *Daucus carota* AB034249, *Croton sublyratus* AB034250, *Scoparia dulcis* AF020041, *Helianthus annuus* AF049658, *Drosophila melanogaster* signal recognition particle 19kDa protein (*srp19*) gene,partial sequence; and geranylgeranyl pyrophosphate synthase (*quemao*) gene,complete cds AF049659, *Drosophila melanogaster* geranylgeranyl pyrophosphate synthase mRNA, complete cds AF139916, *Brevibacterium linens* AF279807, *Penicillium paxilli* geranylgeranyl pyrophosphate synthase (*ggs1*) gene, complete AF279808, *Penicillium paxilli* dimethylallyl tryptophan synthase (*paxD*) gene, partial cds;and cytochrome P450 monooxygenase (*paxQ*), cytochrome P450 monooxygenase (*paxP*),PaxC (*paxC*), monooxygenase (*paxM*), geranylgeranyl pyrophosphate synthase (*paxG*),PaxU (*paxU*), and metabolite transporter (*paxT*) genes, complete cds AJ010302, *Rhodobacter sphaeroides* AJ133724, *Mycobacterium aurum* AJ276129, *Mucor circinelloides f. lusitanicus carG* gene for geranylgeranyl pyrophosphate synthase, exons 1-6 D85029, *Arabidopsis thaliana* mRNA for geranylgeranyl pyrophosphate synthase, partial cds L25813, *Arabidopsis thaliana* L37405, *Streptomyces griseus* geranylgeranyl pyrophosphate synthase (*crtB*), phytoene desaturase (*crtE*) and phytoene synthase (*crtI*) genes, complete cds U15778, *Lupinus albus* geranylgeranyl pyrophosphate synthase (*ggps1*) mRNA, complete cds U44876, *Arabidopsis thaliana* pregeranylgeranyl pyrophosphate synthase (*GGPS2*) mRNA, complete cds X92893, *C. roseus* X95596, *S. griseus*

Gene	***GenBank Accession Number and Source Organism***
	X98795, *S. alba* Y15112, *Paracoccus marcusii*
crtX (Zeaxanthin glucosylase)	D90087, *E. uredovora* M87280 and M90698, *Pantoea agglomerans*
crtY (Lycopene-β-cyclase)	AF139916, *Brevibacterium linens* AF152246, *Citrus* x paradisi AF218415, *Bradyrhizobium sp.* ORS278 AF272737, *Streptomyces griseus* strain IFO13350 AJ133724, *Mycobacterium aurum* AJ250827, *Rhizomucor circinelloides* f. *lusitanicus carRP* gene for lycopene cyclase/phytoene synthase, exons 1-2 AJ276965, *Phycomyces blakesleeanus carRA* gene for phytoene synthase/lycopene cyclase, exons 1-2 D58420, *Agrobacterium aurantiacum* D83513, *Erythrobacter longus* L40176, *Arabidopsis thaliana* lycopene cyclase (*LYC*) mRNA, complete cds M87280, *Pantoea agglomerans* U50738, *Arabodopsis thaliana* lycopene epsilon cyclase mRNA, complete cds U50739, *Arabidosis thaliana* lycopene β cyclase mRNA, complete cds U62808, *Flavobacterium* ATCC21588 X74599, *Synechococcus sp. lcy* gene for lycopene cyclase X81787, *N. tabacum CrtL-1* gene encoding lycopene cyclase X86221, *C. annuum* X86452, *L. esculentum* mRNA for lycopene β-cyclase X95596, *S. griseus* X98796, *N. pseudonarcissus*
crtI (Phytoene desaturase)	AB046992, *Citrus unshiu CitPDS1* mRNA for phytoene desaturase, complete cds AF039585, *Zea mays* phytoene desaturase (*pds1*) gene promoter region and exon 1 AF049356, *Oryza sativa* phytoene desaturase precursor (*Pds*) mRNA, complete cds AF139916, *Brevibacterium linens* AF218415, *Bradyrhizobium sp.* ORS278 AF251014, *Tagetes erecta* AF364515, *Citrus* x paradisi D58420, *Agrobacterium aurantiacum* D83514, *Erythrobacter longus* L16237, *Arabidopsis thaliana* L37405, *Streptomyces griseus* geranylgeranyl

Gene	*G nBank Accession Number and Sourc Organism*
	pyrophosphate synthase (*crtB*), phytoene desaturase (*crtE*) and phytoene synthase (*crtI*) genes, complete cds L39266, *Zea mays* phytoene desaturase (*Pds*) mRNA, complete cds M64704, Soybean phytoene desaturase M88683, *Lycopersicon esculentum* phytoene desaturase (*pds*) mRNA, complete cds S71770, carotenoid gene cluster U37285, *Zea mays* U46919, *Solanum lycopersicum* phytoene desaturase (*Pds*) gene, partial cds U62808, *Flavobacterium* ATCC21588 X55289, *Synechococcus pds* gene for phytoene desaturase X59948, *L. esculentum* X62574, *Synechocystis sp. pds* gene for phytoene desaturase X68058, *C. annuum pds1* mRNA for phytoene desaturase X71023, *Lycopersicon esculentum pds* gene for phytoene desaturase X78271, *L. esculentum* (Ailsa Craig) PDS gene X78434, *P. blakesleeanus* (NRRL1555) carB gene X78815, *N. pseudonarcissus* X86783, *H. pluvialis* Y14807, *Dunaliella bardawil* Y15007, *Xanthophyllomyces dendrorhous* Y15112, *Paracoccus marcusii* Y15114, *Anabaena* PCC7210 *crtP* gene Z11165, *R. capsulatus*
crtB (Phytoene synthase)	AB001284, *Spirulina platensis* AB032797, *Daucus carota PSY* mRNA for phytoene synthase, complete cds AB034704, *Rubrivivax gelatinosus* AB037975, *Citrus unshiu* AF009954, *Arabidopsis thaliana* phytoene synthase (*PSY*) gene, complete cds AF139916, *Brevibacterium linens* AF152892, *Citrus* x paradisi AF218415, *Bradyrhizobium sp.* ORS278 AF220218, *Citrus unshiu* phytoene synthase (*Psy1*) mRNA, complete cds AJ010302, *Rhodobacter* AJ133724, *Mycobacterium aurum* AJ278287, *Phycomyces blakesleeanus carRA* gene for lycopene cyclase/phytoene synthase,

Gene	***GenBank Accession Number and Sourc Organism***
	AJ304825, *Helianthus annuus* mRNA for phytoene synthase (*psy* gene) AJ308385, *Helianthus annuus* mRNA for phytoene synthase (*psy* gene) D58420, *Agrobacterium aurantiacum* L23424, *Lycopersicon esculentum* phytoene synthase (*PSY2*) mRNA, complete cds L25812, *Arabidopsis thaliana* L37405, *Streptomyces griseus* geranylgeranyl pyrophosphate synthase (*crtB*), phytoene desaturase (*crtE*) and phytoene synthase (*crtI*) genes, complete cds M38424, *Pantoea agglomerans* phytoene synthase (*crtE*) gene, complete cds M87280, *Pantoea agglomerans* S71770, Carotenoid gene cluster U32636, *Zea mays* phytoene synthase (*Y1*) gene, complete cds U62808, *Flavobacterium* ATCC21588 U87626, *Rubrivivax gelatinosus* U91900, *Dunaliella bardawil* X52291, *Rhodobacter capsulatus* X60441, *L. esculentum GTom5* gene for phytoene synthase X63873, *Synechococcus* PCC7942 *pys* gene for phytoene synthase X68017, *C. annuum psy1* mRNA for phytoene synthase X69172, *Synechocystis sp. pys* gene for phytoene synthase X78814, *N. pseudonarcissus*
crtZ (β-carotene hydroxylase)	D58420, *Agrobacterium aurantiacum* D58422, *Alcaligenes sp.* D90087, *E. uredovora* M87280, *Pantoea agglomerans* U62808, *Flavobacterium* ATCC21588 Y15112, *Paracoccus marcusii*
crtW (β-carotene ketolase)	AF218415, *Bradyrhizobium sp.* ORS278 D45881, *Haematococcus pluvialis* D58420, *Agrobacterium aurantiacum* D58422, *Alcaligenes sp.* X86782, *H. pluvialis* Y15112, *Paracoccus marcusii*
crtO (β-C4-ketolase)	X86782, H.pluvialis Y15112, Paracoccus marcusii
crtU (β-carotene dehydrogenase)	AF047490, Zea mays AF121947, Arabidopsis thaliana

Gene	***GenBank Accession Number and Source Organism***
dehydrogenase)	AF139916, Brevibacterium linens AF195507, Lycopersicon esculentum AF272737, Streptomyces griseus strain IFO13350 AF372617, Citrus x paradisi AJ133724, Mycobacterium aurum AJ224683, Narcissus pseudonarcissus D26095 and U38550, Anabaena sp. X89897, C.annuum Y15115, Anabaena PCC7210 crtQ gene
crtA (spheroidene monooxygenase)	AJ010302, Rhodobacter sphaeroides Z11165 and X52291, Rhodobacter capsulatus
crtC (hydroxyneurosporene synthase)	AB034704, Rubrivivax gelatinosus AF195122 and AJ010302, Rhodobacter sphaeroides AF287480, Chlorobium tepidum U73944, Rubrivivax gelatinosus X52291 and Z11165, Rhodobacter capsulatus Z21955, M.xanthus
crtD (carotenoid 3,4-desaturase)	AJ010302 and X63204, Rhodobacter sphaeroides U73944, Rubrivivax gelatinosus X52291and Z11165, Rhodobacter capsulatus
crtF (1-OH-carotenoid methylase)	AB034704, Rubrivivax gelatinosus AF288602, Chloroflexus aurantiacus AJ010302, Rhodobacter sphaeroides X52291 and Z11165, Rhodobacter capsulatus

The majority of the most preferred *crt* genes are primarily from *Panteoa stewartii*. Sequences of these preferred genes are presented as the following SEQ ID numbers: the *crtE* gene (SEQ ID NO:1), the *crtX* gene (SEQ ID NO:3), *crtY* (SEQ ID NO:5), the *crtI* gene (SEQ ID NO:7), the *crtB* gene (SEQ ID NO:9) and the *crtZ* gene (SEQ ID NO:11).

By using various combinations of the genes presented in Table 2 and the preferred genes of the present invention, innumerable different carotenoids and carotenoid derivatives can be made using the methods of the present invention, provided that sufficient sources of FPP are available in the host organism. For example, the gene cluster *crtEXYIB* enables the production of β-carotene. Addition of the *crtZ* to *crtEXYIB* enables the production of zeaxanthin.

It is envisioned that useful products of the present invention will include any carotenoid compound as defined herein including, but not limited to antheraxanthin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β-cryptoxanthin, didehydrolycopene, didehydrolycopene, β-

carotene, ζ–carotene, δ–carotene, γ–carotene, keto–γ–carotene, ψ–carotene, ε–carotene, β,ψ–carotene, torulene, echinenone, gamma-carotene, zeta-carotene, alpha-cryptoxanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, β-isorenieratene lactucaxanthin, lutein, lycopene, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, uriolide, uriolide acetate, violaxanthin, zeaxanthin-β-diglucoside, zeaxanthin, and C30-carotenoids. Additionally, the invention encompasses derivitization of these molecules to create hydroxy-, methoxy-, oxo-, epoxy-, carboxy-, or aldehydic functional groups, or glycoside esters, or sulfates.

Description of the Preferred Embodiments

Publicly available sequences for several isoprenoid pathway genes in *E. coli* were used to synthesize integration cassettes for λ-Red mediated homologous recombination. One or two PCR-generated fragments were prepared and engineered to contain the phage *T5* promoter and a selection marker (Example 1, Tables 1-3). Homology arms, approximately 40-50 bp in length were used on the ends of the PCR generated fragment(s). Homologous recombination, aided by the λ-Red recombinase system encoded on plasmid pKD46, occurred between the *E. coli* chromosome and the integration cassettes, effectively replacing the native promoter of the *dxs, idi, lytB, dxr, ygbBygbP(ygbBP), ispA, ychB, gcpE,* and *ispB* genes with the P_{T5} strong promoter (Figures 1, 3, and 5). Colonies of transformants were obtained for each of these genes. Successful recombination was measured by the inclusion of a selectable marker (kanamycin). Chromosomal integration of the integration cassettes was confirmed via PCR analysis as described in Example 1 (Figure 8). λ-Red mediated recombination can occur with one or more integration cassettes, however, the use of at least 2 linear, PCR-generated, cassettes is preferred (Figure 3).

In another embodiment, a reporter strain of *E. coli* was constructed for assaying β-carotene production. Briefly, the *E. coli* reporter strain was created by cloning the gene cluster *crtEXYIB* from *Pantoea stewartii* into a helper plasmid (pPCB15; SEQ ID NO:40), which was subsequently used to transform the *E. coli* host strain (Figure 6). The cluster contained many of the genes required for the synthesis of carotenoids (i.e. β-carotene). It should be noted that the *crtZ* gene (β-carotene hydroxylase) was included

in the gene cluster. However, since no promoter was present to express the *crt*Z gene (organized in opposite orientation and adjacent to *crt*B gene) no zeaxanthin was produced, thus, the zeaxanthin glucosyl transferase enzyme (encoded by the *crt*X gene located within the gene cluster) had no substrate for its reaction. Increases in β-carotene production were reported as increases relative to the control strain production (Figure 6).

Sequence analysis was conducted to confirm the identification of the gene cluster (Example 4, Table 4) In order to confirm the function of the various carotenoid genes on the helper plasmid, transposon mutagenesis (Example 5) was used. Using this method it was possible to assign function to each of the genes from the transposon mutagenesis results (Table 5). The function assigned to the various *Pantoea stewartii crt* genes was in agreement with that reported in the art (WO 02/079395 A2 and WO 03/016503).

The reporter plasmid, pPCB15, was used to monitor increased flux through the isoprenoid pathway. Modifications to the isoprenoid pathway, which altered the amount of FPP produced, were monitored by the production of β-carotene. Comparisons between the amount of β-carotene produced in the wild type *E. coli* strain and the various transformants were used to select for those strains exhibiting optimal β-carotene production.

In another embodiment, the present method illustrates the ability to use P1 transduction to create an *E. coli* strains having increased β-carotene production. The *E. coli kan-*P_{T5}*-dxs* strain created in Example 1 was infected with bacteriophage P1. Lysate was collected and used to infect *E. coli* MG1655 containing the β-carotene expression plasmid pPCB15 (Example 6, Figure 6). Transductants were selected via the kanamycin selection marker. A temperature sensitive helper plasmid (pCP20), encoding a site-specific recombinase (FLP), was used to remove the selection marker (ATCC PTA-4455; Cherepanov and Wackernagel, *supra*; Example 6). The plasmid was cured after removal of the selection marker. PCR fragment analysis was used to confirm incorporation of the P_{T5}*-dxs* cassette and removal of the kanamycin marker (Figure 8). *E. coli* P_{T5}*-dxs* exhibited approximately a 2.8 fold increase in β-carotene production in comparison to the wild type *E. coli* containing the reporter plasmid (Example 8, Figure 9).

In order to optimize flux through the isoprenoid pathway, transformants were created containing multiple chromosomal

modifications. Bacteriophage P1 transduction was used to create the multiple transformants. Each of the strains were prepared as described in Example 1 and infected with bacteriophage P1. The lytic cycle was allowed to proceed. The lysates of each strain were collected. A mixture of P1 lysates was prepared by mixing equal titers of P1 lysates from each of the individual strains (Example 2, Figure 7). The P1 lysate mixture was used to infect the *E. coli* P_{T5}-*dxs* strain. Transductants were selected via the selection marker. Transductants exhibiting the deeper yellow pigmentation were selected. Once again, the selection marker was removed. PCR fragment analysis was performed to identify the location and type of insertion for each of the selected transductants (Example 9) and to identify removal of the selection marker (Figure 8). Removal of the selection marker allowed for the selected strains to act as recipient cells for the next round of P1 transduction. The process was repeated, creating an optimized *E. coli* strain P_{T5}-*dxs* P_{T5}-*idi* P_{T5}-*ispB* exhibiting a 3.4-fold increase in β-carotene production in comparison to the control strain (Figure 9). Using this process it was possible to efficiently engineer multiple chromosomal modifications into *E.coli*. The isoprenoid pathway was optimized to increase the production of β-carotene, the genetic end product of interest.

In another embodiment, the inclusion of *ispB* as one of the targets for up-regulation was unexpected as it was believe to divert the carbon flow from the isoprenoid pathway (Figures 1 and 9; Example 9). The present method allowed for identification of gene and gene combinations that may be altered using the present method to increased production β-carotene.

In another embodiment, the integration cassettes used in the present method may contain disrupted genes, such as those disrupted by transposon mutagenesis. Down-regulating or completely disrupting genes via chromosomal engineering allows one to divert carbon flow away of competing biosynthetic pathways. The present method facilitates assessment of various combinations of chromosomal modifications and their effect on the desired genetic end product of the targeted biosynthetic pathway.

In one embodiment, the bacterial host strain is engineered to contain multiple chromosomal modifications, including multiple promoter replacements so that the production efficiency of the desired genetic end product is increased. Multiple chromosomal modifications were integrated

into one host strain using P1 transduction and a site-specific recombinase to remove selectable markers. Chromosomal modifications were integrated successively into a single strain by successive rounds of P1 transduction and marker removal.

The invention may be used for stacking a variety of targeted *in vivo* bacterial chromosomal modifications into a single host strain. The removal of the selectable marker using a site-specific recombinase allows for one to conduct multiple chromosomal modifications, necessary for engineering biosynthetic pathways and for optimizing production of industrially useful materials. A combinatorial approach to stacking traits allows the integration of chromosomal modifications with the most impact on the desired trait to be obtained more quickly in comparison to assessing the impact of individual modifications one at a time.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, (1989) (Maniatis) and by T. J. Silhavy, M. L. Bennan, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987).

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994))

or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

Manipulations of genetic sequences were accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). Where the GCG program "Pileup" was used the gap creation default value of 12, and the gap extension default value of 4 were used. Where the CGC "Gap" or "Bestfit" programs were used the default gap creation penalty of 50 and the default gap extension penalty of 3 were used. Multiple alignments were created using the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-120. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY). In any case where program parameters were not prompted for, in these or any other programs, default values were used.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "mL" means milliliters, "L" means liters, "μL" means microliters, and "rpm" means revolutions per minute.

EXAMPLE 1

Synthesis of *E. coli* Strains with the Phage *T5* Strong Promoter Chromosomally Integrated Upstream of the Isoprenoid Genes (Promoter Replacement)

The native promoters of the *E. coli* isoprenoid genes, *dxs, idi, dxr, lytB, ygbBygbP(ygbBP), ispA, ychB, gcpE,* and *ispB* (Figure 1) were replaced with the phage *T5* (P_{T5}) strong promoter using a PCR-fragment chromosomal integration method as described in Figure 3. The method for replacement is based on homologous recombination via the λ Red recombinase encoded on a helper plasmid. Recombination occurs between the *E. coli* chromosome and one or two PCR fragments that contain 40-50 bp homology patches at both ends of PCR fragments (Figure 3). Either a two PCR fragment or one PCR fragment method (Figure 3) was used for chromosomal integration of the kanamycin

selectable marker and phage *T5* promoter (SEQ ID NO:43) in the front of the *E. coli* isoprenoid genes, *dxs, idi, lytB, dxr, ygbBygbP(ygbBP), ispA, ychB, gcpE*, and *ispB*. For the two PCR fragment method, the two fragments included a linear DNA fragment (1489 bp) containing a kanamycin selectable marker flanked by site-specific recombinase target sequences (*FRT*) and a linear DNA fragment (154 bp) containing a phage *T5* promoter (P_{T5}) comprising the –10 and –35 consensus promoter sequences, lac operator (*lacO*), and a ribosomal binding site (*rbs*). For the one PCR fragment method, the fused linear DNA fragment (1647 bp) contained a kanamycin selectable marker flanked by site-specific recombinase target sequences (*FRT*) and a linear DNA fragment (154 bp) containing a P_{T5} promoter comprising the –10 and –35 consensus promoter sequences, lac operator (*lacO*), and a ribosomal binding site (*rbs*).

By using the two PCR fragment method, the kanamycin selectable marker and P_{T5} promoter (*kan-*P_{T5}) were integrated upstream of the *dxs, idi, lytB, dxr*, and *ygbBP* genes, replacing the native promoter of each, yielding *kan-*P_{T5}*-dxs, kan-*P_{T5}*-idi, kan-*P_{T5}*-lytB, kan-*P_{T5}*-dxr, and kan-*P_{T5}*-ygbBP*. The linear DNA fragment (1489 bp) which contained a kanamycin selectable marker was synthesized by PCR from plasmid pKD4 (Datsenko and Wanner, *supra*) with primer pairs as follows in Table 3.

TABLE 3

Primers for Amplification of the Kanamycin Selectable Marker

Primer Name	Primer Sequence	SEQ ID NO:
5'-kan(dxs)	TGGAAGCGCTAGCGGACTACATCATCCAG CGTAATAAATAACGTCTTGAGCGATTGTGT AG[1]	13
5'-kan(idi)	TCTGATGCGCAAGCTGAAGAAAAATGAGC ATGGAGAATAATATGACGTCTTGAGCGATT GTGTAG[1]	14
5'-kan(lytB)	TTTGATATTGAAGTGCTGGAAATCGATCCG GCACTGGAGGCGTAACGTCTTGAGCGATT GTGTAG[1]	15
5'-kan(dxr)	GAAGCGGCGCTGGCAGACAAAGAAGCAG AACTGATGCAGTTCTGACGTCTTGAGCGAT TGTGTAG[1]	16
5'-kan(ygbBP)	GACGCGTCGAAGCGCGCACAGTCTGCGG GGCAAAACAATCGATAACGTCTTGAGCGA	17

Primer Name	Primer Sequence	SEQ ID NO:
	TTGTGTAG[1]	
3'-kan	GAAGACGAAAGGGCCTCGTGATACGCCTA TTTTTATAGGTTATATGAATATCCTCCTTAG TTCC[2]	18

[1] The underlined sequences illustrate each respective homology arm chosen to match sequences in the upstream region of the chromosomal integration site, while the remainder is the priming sequence.
[2] The underlined sequences illustrate homology arm chosen to match sequences in the 5'-end region of the T5 promoter DNA fragment.

The second linear DNA fragment (154 bp) containing a P_{T5} promoter was synthesized by PCR from pQE30 (QIAGEN, Inc. Valencia, CA) with primer pairs as follows in Table 4.

TABLE 4
Primers for Amplification of the *T5* Promoter

Primer Name	Primer Sequence	SEQ ID NO:
5'-T5	CTAAGGAGGATATTCATATAACCTATAAAAA TAGGCGTATCACGAGGCCC[1]	19
3'- T5(dxs)	GGAGTCGACCAGTGCCAGGGTCGGGTATT TGGCAATATCAAAACTCATAGTTAATTTCTC CTCTTTAATG[2]	20
3'- T5(idi)	TGGGAACTCCCTGTGCATTCAATAAAATGA CGTGTTCCGTTTGCATAGTTAATTTCTCCTC TTTAATG[2]	21
3'- T5(lytB)	CTACCCCGGCACAAAAACCACGTGGGTTG GCCAACAGGATCTGCATAGTTAATTTCTCCT CTTTAATG[2]	22
3'- T5(dxr)	TGCAACCAATCGAGCCGGTCGAGCCCAGA ATGGTGAGTTGCTTCATAGTTAATTTCTCCT CTTTAATG[2]	23
3'- T5(ygbBP)	CGGCCGCCGGAACCACGGCGCAAACATCC AAATGAGTGGTTGCCATAGTTAATTTCTCCT CTTTAATG[2]	24

[1] The underlined sequences illustrate homology arm chosen to match sequences in the 3'-end region of the kanamycin DNA fragment.
[2] The underlined sequences illustrate each respective homology arm chosen to match sequences in the downstream region of the chromosomal integration site.

For the one PCR fragment method, the kanamycin selectable marker and phage *T5* promoter were integrated in the front of *ispA, ychB, gcpE*, and *ispB* genes, yielding *kan-*P_{T5}*-ispA, kan-*P_{T5}*- ychB, kan-*P_{T5}*-*

gcpE, and *kan-P_{T5}-ispB*. The linear DNA fragment used for integration and which contained a fused kanamycin selectable marker-phage with P_{T5} promoter was synthesized by PCR from pSUH5 (Figure 4) with primer pairs as follows in Table 5.

TABLE 5

Primers for Amplification of the Fragment for the One PCR Fragment Method

Primer Name	Primer Sequence	SEQ ID NO:
5'-kanT5(ispA)	AACGAAGACGCCTCTCTAACCCCTTTTACACCGGACAATGAGTAACGTCTTGAGCGATTGTGTAG[1]	25
5'-kanT5(ychB)	GGTCAACGCATCAAGTTAAAAATGGATAACTGGATAGTGAAATAACGTCTTGAGCGATTGTGTAG[1]	26
5'-kanT5(gcpE)	GTTGCGCGTCTGACCCTCAATGCCGAACAATCACCGGCGCAGTAACGTCTTGAGCGATTGTGTAG[1]	27
5'-kanT5(ispB)	ACCATAAACCCTAAGTTGCCTTTGTTCACAGTAAGGTAATCGGGGCGTCTTGAGCGATTGTGTAG[1]	28
3'-kanT5(ispA)	CTGGTTGGCCTGCTTAACGCAGGCTTCGAGTTGCTGCGGAAAGTCCATAGTTAATTTCTCCTCTTTAATG[2]	29
3'-kanT5(ychB)	ATAAAAACAGATTAAGTTTTGCCGGAGAGGGCCACTGTGTCCGCATAGTTAATTTCTCCTCTTTAATG[2]	30
3'-kanT5(gcpE)	AAATACGTGTTGATTTTCTACGTTGAATTGGAGCCTGGTTATGCATAGTTAATTTCTCCTCTTTAATG[2]	31
3'-kanT5(ispB)	CGCCATATCTTGCGCGGTTAACTCATTGATTTTTTCTAAATTCATAGTTAATTTCTCCTCTTTAATG[2]	32

[1]The underlined sequences illustrate each respective homology arm chosen to match sequences in the upstream region of the chromosomal integration site.
[2]The underlined sequences illustrate each respective homology arm chosen to match sequences in the downstream region of the chromosomal integration site.

Standard PCR conditions were used to amplify the linear DNA fragments with Ampli*Taq* Gold® polymerase (Applied Biosystems, Foster City, CA) as follows:

PCR reaction:	PCR reaction mixture:
Step1 94°C 3 min	0.5 μL plasmid DNA
Step2 93°C 30 sec	5 μL 10X PCR buffer
Step3 55°C 1 min	1 μL dNTP mixture (10 mM)
Step4 72°C 3 min	1 μL 5'-primer (20 μM)
Step5 Go To Step2, 30 cycles	1 μL 3'-primer (20 μM)
Step6 72°C 5 min	0.5 μL Ampli*Taq* Gold® polymerase
	41 μL sterilized dH_2O

After completing the PCR reactions, 50 μL of each PCR reaction mixture was run on a 1% agarose gel and the PCR products were purified using the QIAquick Gel Extraction Kit™ as per the manufacturer's instructions (Cat. # 28704, QIAGEN Inc., Valencia, CA). The PCR products were eluted with 10 μL of distilled water. The DNA Clean & Concentrator™ kit (Zymo Research, Orange, CA) was used to further purify the PCR product fragments as per the manufacturer's instructions. The PCR products were eluted with 6-8 μL of distilled water to a concentration of 0.5-1.0 μg/μL.

The *E. coli* MC1061 strain, carrying a λ-Red recombinase expression plasmid pKD46 (amp^R) (Figure 5), was used as a host strain for the chromosomal integration of the PCR fragments. The strain was constructed by transformation of *E. coli* strain MC1061 with the λ-Red recombinase expression plasmid, pKD46 (amp^R). The λ-Red recombinase in pKD46 is comprised of three genes *exo, bet,* and *gam* expressed under the control of an arabinose-inducible promoter. Transformants were selected on 100 μg/mL ampicillin LB plates at 30°C.

For transformation, electroporation was performed using 1-5 μg of the purified PCR products carrying the kanamycin marker and P_{T5} promoter. Approximately one-half of the cells transformed were spread on LB plates containing 25 μg/mL kanamycin in order to select antibiotic resistant transformants. After incubating the plate at 37°C overnight, antibiotic-resistant transformants were selected as follows: 10 colonies of *kan-P_{T5}-dxs,* 12 colonies of *kan-P_{T5}-idi,* 1 colony of *kan-P_{T5}-lytB,* 47 colonies of *kan-P_{T5}-dxr,* 10 colonies of *kan-P_{T5}-ygbBP,* 19 colonies of *kan-P_{T5}-ispA,* 700 colonies of *kan-P_{T5}-ychB,* 21 colonies of *kan-P_{T5}-gcpE,* and 3 colonies of *kan-P_{T5}-ispB.*

PCR analysis was used to screen the selected *kan-P_{T5}* kanamycin-resistant transformants for integration of both the kanamycin selectable

marker and the phage *T5* promoter (P_{T5}) in the correct location on the *E. coli* chromosome. For PCR, a colony was resuspended in 50 μL of PCR reaction mixture containing 200 μM dNTPs, 2.5 U Ampli*Taq*™ (Applied Biosytems), and 0.4 μM of specific primer pairs. Test primers were chosen to match sequences of the regions located in the kanamycin (5'-primer) and the early coding-region of each isoprenoid gene (3'-primer). The PCR reaction was performed as described in above. Chromosomal integration of *kan-*P_{T5} upstream of each isoprenoid gene was confirmed by PCR analysis. The resultant *E. coli* strains carrying each *kan-*P_{T5}-isoprenoid gene fusions on the chromosome were used for stacking multiple *kan-*P_{T5}-isoprenoid gene fusions in parallel on the chromosome in a combinatorial approach as described in Examples 7 and 9.

EXAMPLE 2

Preparation of P1 Lysate Mixture Made with the *E. coli kan-*P_{T5}*-dxs, kan-*P_{T5}*-idi, kan-*P_{T5}*-lytB, kan-*P_{T5}*-dxr, kan-*P_{T5}*-ygbBP, kan-*P_{T5}*-ispA, kan-*P_{T5}*-ychB, kan-*P_{T5}*-gcpE,* and *kan-*P_{T5}*-ispB* Strains

P1 lysates of the *E. coli kan-*P_{T5}*-dxs, kan-*P_{T5}*-idi, kan-*P_{T5}*-lytB, kan-*P_{T5}*-dxr, kan-*P_{T5}*-ygbBP, kan-*P_{T5}*-ispA, kan-*P_{T5}*-ychB, kan-*P_{T5}*-gcpE,* and *kan-*P_{T5}*-ispB* strains were prepared by infecting a growing culture of bacteria with the P1 phage and allowing the cells to lyse. For P1 infection, each strain was inoculated in 4 mL LB medium with 25 μg/mL kanamycin, grown at 37°C overnight, and then sub-cultured with 1:100 dilution of an overnight culture in 10 mL LB medium containing 5 mM $CaCl_2$. After 20-30 min of growth at 37°C, 10^7 $P1_{vir}$ phages were added. The cell-phage mixture was aerated for 2-3 hr at 37°C until lysed, several drops of chloroform were added and the mixture vortexed for 30 sec and incubated for an additional 30 min at room temp. The mixture was then centrifuged for 10 min at 4500 rpm, and the supernatant transferred into a new tube to which several drops of chloroform were added. The lysates were stored at 4°C.

A mixture of P1 lysates was prepared by mixing equal titers of P1 lysate from *E. coli kan-*P_{T5}*-dxs, kan-*P_{T5}*-idi, kan-*P_{T5}*-lytB, kan-*P_{T5}*-dxr, kan-*P_{T5}*-ygbBP, kan-*P_{T5}*-ispA, kan-*P_{T5}*-ychB, kan-*P_{T5}*-gcpE,* or *kan-*P_{T5}*-ispB* strains. Titer measurements of the P1 lysates were determined as described in Maniatis.

EXAMPLE 3

Cloning of β-Carotene Production Genes from *Pantoea stewartii*

Primers were designed using the sequence from *Erwinia uredovora* to amplify a fragment by PCR containing the *crt* genes. These sequences included 5'-3':

ATGACGGTCTGCGCAAAAAAACACG	SEQ ID 33
GAGAAATTATGTTGTGGATTTGGAATGC	SEQ ID 34

Chromosomal DNA was purified from *Pantoea stewartii* (ATCC no. 8199) and *Pfu* Turbo polymerase (Stratagene, La Jolla, CA) was used in a PCR amplification reaction under the following conditions: 94°C, 5 min; 94°C (1 min)-60°C (1 min)-72°C (10 min) for 25 cycles, and 72°C for 10 min. A single product of approximately 6.5-kb was observed following gel electrophoresis. *Taq* polymerase (Perkin Elmer, Foster City, CA) was used in a ten minute 72°C reaction to add additional 3' adenosine nucleotides to the fragment for TOPO cloning into pCR4-TOPO (Invitrogen, Carlsbad, CA) to create the plasmid pPCB13. Following transformation to *E. coli* DH5α (Life Technologies, Rockville, MD) by electroporation, several colonies appeared to be bright yellow in color indicating that they were producing a carotenoid compound. Following plasmid isolation as instructed by the manufacturer using the Qiagen (Valencia, CA) miniprep kit, the plasmid containing the 6.5 kb amplified fragment was transposed with pGPS1.1 using the GPS-1 Genome Priming System kit (New England Biolabs, Inc., Beverly, MA). A number of these transposed plasmids were sequenced from each end of the transposon. Sequence was generated on an ABI Automatic sequencer using dye terminator technology (US 5,366,860; EP 272007) using transposon specific primers. Sequence assembly was performed with the Sequencher program (Gene Codes Corp., Ann Arbor MI).

EXAMPLE 4

Identification and Characterization of Bacterial Genes

Genes encoding *crtE, X, Y, I, B*, and *Z* were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank® CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The

sequences obtained in Example 3 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J., *Nature Genetics,* 3:266-272 (1993)) provided by the NCBI.

All comparisons were done using either the BLASTNnr or BLASTXnr algorithm. The results of the BLAST comparison are given in Table 6, which summarize the sequences to which they have the most similarity. Table 6 displays data based on the BLASTXnr algorithm with values reported in expect values. The Expected value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

TABLE 6

ORF Name	Gene Name	Similarity Identified	SEQ ID No. Nucleotide	SEQ ID No. Peptide	% Identity a	% Similarity b	E-value[c]	Citation
1	*crtE*	Geranylgeranyl pryophosphate synthetase (or GGPP synthetase, or farnesyltranstransferase) EC 2.5.1.29 gi|117509|sp|P21684|CRTE_PANAN GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE) (FARNESYLTRANSTRANSFERASE)	1	2	83	88	e-137	Misawa et al., *J. Bacteriol.* 172 (12), 6704-6712 (1990)
2	*crtX*	Zeaxanthin glucosyl transferase EC 2.4.1.- gi|1073294|pir||S52583 crtX protein - *Erwinia herbicola*	3	4	75	79	0.0	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)
3	*crtY*	Lycopene cyclase gi|1073295|pir||S52585 lycopene cyclase - *Erwinia herbicola*	5	6	83	91	0.0	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)
4	*crtI*	Phytoene desaturaseEC 1.3.-.- gi|1073299|pir||S52586 phytoene dehydrogenase (EC 1.3.-.-) - *Erwinia herbicola*	7	8	89	91	0.0	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)

5	*crtB*	Phytoene synthaseEC2.5.1.- gi|1073300|pir||S52587 prephytoene pyrophosphate synthase - *Erwinia herbicola*	9	10	88	92	e-150	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)
6	*crtZ*	Beta-carotene hydroxylase gi|117526|sp|P21688|CRTZ_PANAN BETA-CAROTENE HYDROXYLASE	11	12	88	91	3e-88	Misawa et al., *J. Bacteriol.* 172 (12), 6704-6712 (1990)

[a]%Identity is defined as percentage of amino acids that are identical between the two proteins.
[b]% Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.
[c]Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 5

Analysis of Gene Function by Transposon Mutagenesis

Several plasmids carrying transposons, which were inserted into each coding region including *crtE*, *crtX*, *crtY*, *crtI*, *crtB*, and *crtZ*, were chosen using sequence data generated in Example 3. These plasmid variants were transformed to *E. coli* MG1655 and grown in 100 mL Luria-Bertani broth in the presence of 100 μg/mL ampicillin. Cultures were grown for 18 hr at 26°C, and the cells were harvested by centrifugation. Carotenoids were extracted from the cell pellets using 10 mL of acetone. The acetone was dried under nitrogen and the carotenoids were resuspended in 1 mL of methanol for HPLC analysis. A Beckman System Gold® HPLC with Beckman Gold Nouveau Software (Columbia, MD) was used for the study. The crude extraction (0.1 mL) was loaded onto a 125 x 4 mm RP8 (5 μm particles) column with corresponding guard column (Hewlett-Packard, San Fernando, CA). The flow rate was 1 mL/min, while the solvent program used was: 0-11.5 min 40% water/60% methanol; 11.5-20 min 100% methanol; 20-30 min 40% water/60% methanol. The spectrum data were collected by the Beckman photodiode array detector (model 168).

In the wild type clone with wild type *crtEXYIBZ*, the carotenoid was found to have a retention time of 15.8 min and an absorption spectra of 425, 450 and 478 nm. These values matched those of the β-carotene standard. This suggested that *crtZ* gene organized in the opposite orientation was not expressed in this construct. The transposon insertion in *crtZ* had no effect as expected (data not shown).

HPLC spectral analysis also revealed that a clone with transposon insertion in *crtX* also produced β-carotene. This is consistent with the proposed function of *crtX* encoding a zeaxanthin glucosyl transferase enzyme at a later step of the carotenoid pathway following synthesis of β-carotene.

The transposon insertion in *crtY* did not produce β-carotene. The carotenoid's elution time (15.2 min) and absorption spectra (443 nm, 469 nm, 500 nm) agreed with those of the lycopene standard. Accumulation of lycopene in the *crtY* mutant confirmed the role *crtY* as a lycopene cyclase encoding gene.

The *crtI* extraction, when monitored at 286 nm, had a peak with retention time of 16.3 min and with absorption spectra of 276 nm, 286 nm, 297 nm, which agreed with the reported spectrum for phytoene. Detection

of phytoene in the *crtI* mutant confirmed the function of the *crtI* gene as one encoding a phytoene dehydrogenase enzyme.

The acetone extracted from the *crtE* mutant or *crtB* mutant was clear. Loss of pigmented carotenoids in these mutants indicated that both the *crtE* gene and *crtB* genes are essential for carotenoid synthesis. No carotenoid was observed in either mutant, which is consistent with the proposed function of *crtB* encoding a prephytoene pyrophosphate synthase and *crtE* encoding a geranylgeranyl pyrophosphate synthetase. Both enzymes are required for β-carotene synthesis.

Results of the transposon mutagenesis experiments are shown below in Table 7. The site of transposon insertion into the gene cluster *crtEXYIB* is recorded, along with the color of the *E. coli* colonies observed on LB plates, the identity of the carotenoid compound (as determined by HPLC spectral analysis), and the experimentally assigned function of each gene.

Table 7
Transposon Insertion Analysis of Carotenoid Gene Function

Transposon insertion site	Colony color	Carotenoid observed by HPLC	Assigned gene function
Wild Type (with no transposon insertion)	Yellow	β-carotene	
crtE	White	None	Geranylgeranyl pyrophosphate synthetase
crtB	White	None	Prephytoene pyrophosphate synthase
crtI	White	Phytoene	Phytoene dehydrogenase
crtY	Pink	Lycopene	Lycopene cyclase
crtZ	Yellow	β-carotene	β-carotene hydroxylase
crtX	Yellow	β-carotene	Zeaxanthin glucosyl transferase

EXAMPLE 6

Construction of *E. coli* P_{T5}-*dxs* that Produces β-Carotene

In order to characterize the effect of the phage *T5* promoter on isoprenoid production, a strain, *E. coli* P_{T5}-*dxs*, containing a

chromosomally integrated *T5* promoter upstream from an isoprenoid gene, capable of producing β-carotene, was constructed.

P1 lysate made on *E. coli kan-P_{T5}-dxs* strain was transduced into the recipient strain, *E. coli* MG1655 containing a β-carotene biosynthesis expression plasmid pPCB15 (cam^R) (Figure 6). The plasmid pPCB15 (cam^R) contains the carotenoid biosynthesis gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC no. 8199). The pPCB15 plasmid was constructed from ligation of *Sma*I digested pSU18 (Bartolome, B. et al., *Gene*, 102:75-78 (1991)) vector with a blunt-ended *PmeI/NotI* fragment carrying *crtEXYIB* from pPCB13 (Example 3). The *E. coli* MG1655 pPCB15 recipient cells were grown to mid-log phase (1-2 x 10^8 cells/mL) in 4 mL LB medium with 25 μg/mL chloramphenicol at 37°C. Cells were spun down for 10 min at 4500 rpm and resuspended in 2 mL of 10 mM $MgSO_4$ and 5 mM $CaCl_2$. Recipient cells (100 μL) were mixed with 1 μL, 2 μL, 5 μL, or 10 μL of P1 lysate stock (10^7 pfu/μL) made from the *E. coli kan-P_{T5}-dxs* strain and incubated at 30°C for 30 min. The recipient cell-lysate mixture was spun down at 6500 rpm for 30 sec, resuspended in 100 μL of LB medium with 10 mM of sodium citrate, and incubated at 37°C for 1 h. Cells were plated on LB plates containing both 25 μg/mL kanamycin and 25 μg/mL of chloramphenicol in order to select for antibiotic-resistant transductants and incubated at 37°C for 1 or 2 days. Sixteen transductants were selected.

To eliminate kanamycin selectable marker from the chromosome, a FLP recombinase expression plasmid pCP20 (amp^R) (ATCC PTA-4455) (Cherepanov and Wackernagel, *Gene*, 158:9-14 (1995)), which has a temperature-sensitive replication of origin, was transiently transformed into one of the kanamycin-resistant transductants by electroporation. Cells were spread onto LB agar containing 100 μg/mL ampicillin and 25 μg/mL chloramphenicol LB plates, and grown at 30°C for 1 day. Colonies were picked and streaked on 25 μg/mL chloramphenicol LB plates without ampicillin antibiotics and incubated at 43°C overnight. Plasmid pCP20 has a temperature sensitive origin of replication and was cured from the host cells by culturing cells at 43°C. The colonies were tested for ampicillin and kanamycin sensitivity to test loss of pCP20 and kanamycin selectable marker by streaking colonies on 100 μg/mL ampicillin LB plate or 25 μg/mL kanamycin LB plate. Elimination of the kanamycin selectable marker from the *E. coli* chromosome was confirmed by PCR analysis (Figure 8, lane 1 and 2). The selected colonies were resuspended in 50

μL of PCR reaction mixture containing 200 μM dNTPs, 2.5 U Ampli*Taq*™ (Applied Biosytems), and 0.4 μM of different combination of specific primer pairs, T-kan (5'-ACCGGATATCACCACTTAT CTGCTC-3')(SEQ ID NO:35) and B-dxs (5'-TGGCAACAGTCGTAGCTCCTGGG TGG-3')(SEQ ID NO:36), T-T5 (5'-TAACCTATAAAAATAGGCGTATCACGAGG CCC-3')(SEQ ID NO:37) and B-dxs. Test primers were chosen to amplify regions located either in the kanamycin or the phage T5 promoter and the early region of *dxs* gene (Figure 8). The PCR reaction was performed as described in Example 1. The PCR results (Figure 8, lane 1 and 2) indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome. The presence of the P_{T5} promoter fragment upstream of the *dxs* coding sequence was confirmed based on the production of a PCR product of the expected size (229 bp). In this manner the *E. coli* P_{T5}-*dxs* strain was constructed.

EXAMPLE 7

Combinatorial Stacking of Multiple *kan-*P_{T5}-Isoprenoid Gene Fusions in Parallel

In order to create a bacterial strain capable of high level of carotenoid production, a method was devised for stacking P_{T5} in front of multiple isoprenoid genes in parallel. Using this technique enabled one to then select for the best carotenoid producing strain.

E. coli P_{T5}-*dxs* strain was transduced with P1 lysate mixture made with *E. coli kan-*P_{T5}*-idi, kan-*P_{T5}*-lytB, kan-*P_{T5}*-dxr, kan-*P_{T5}*-ygbBP, kan-*P_{T5}*-ispA, kan-*P_{T5}*-ychB, kan-*P_{T5}*-gcpE,* and *kan-*P_{T5}*-ispB* strains as described in Example 2, which allowed stacking *kan-*P_{T5} cassettes in front of multiple isoprenoid genes in parallel (Figure 7). For transduction, the recipient cells were prepared and transduction was carried out as in Example 6 using the P1 lysate mixture (10^6 pfu/μL). Cells were plated on LB plates containing both 25 μg/mL kanamycin and 25 μg/mL chloramphenicol in order to select antibiotic-resistant transductants. After incubation at 37°C for 1-2 days, six colonies out of 430 kanamycin/chloramphenicol-resistant transductants that were most deeply pigmented with the characteristic yellow β-carotene color were selected. The *kan-*P_{T5}-isoprenoid gene fusions stacked on the chromosome in these six strains were identified by PCR analysis with a 5'-primer complementary to the middle region of the kanamycin gene and a 3'-primer complementary to the sequence within the first several hundred bp of each isoprenoid gene (*idi, lytB, dxr, ygbBP, ispA, ychB, gcpE,* or *ispB*).

This PCR screening was performed as outlined in Example 1. PCR analysis showed that in addition to the *P_{T5}-dxs,* four colonies contained *kan-P_{T5}-idi,* one contained *kan-P_{T5}-ispB,* and one contained *kan-P_{T5}-gcpE.* Among these, colonies carrying *kan-P_{T5}-idi* showed the deepest yellow color on an LB plate containing both 25 μg/mL kanamycin and 25 μg/mL chloramphenicol after growth at 37°C for 2 days, which suggested higher yields of β-carotene production.

The kanamycin selectable marker from the chromosome of *E. coli P_{T5}-dxs kan-P_{T5}-idi* was eliminated as described in Example 2, yielding *E. coli P_{T5}-dxs P_{T5}-idi.* The elimination of the kanamycin selectable marker was confirmed by PCR analysis as described in Example 1 using different combinations of specific primer pairs, T-kan and B-idi (5'-TCATGCTGACCTGGTGAAGGAATCC-3')(SEQ ID NO:38), T-T5 and B-idi. Test primers were chosen to amplify regions located either in the kanamycin or the P_{T5} promoter and the beginning of the *idi* gene (Figure 8). The PCR results (Figure 8, lane 3 and 4) indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome. As before, the presence of the P_{T5} promoter fragment in the front of chromosomal *idi* gene was confirmed based on the production of a PCR fragment of the expected size (274 bp).

EXAMPLE 8

Measurement of β-carotene Production in *E. coli P_{T5}-dxs P_{T5}-idi*

β-carotene production of *E. coli P_{T5}-dxs P_{T5}-idi*, *E. coli P_{T5}-dxs*, and *E. coli* control strains all of which contain a β-carotene biosynthesis expression plasmid pPCB15 (cam^R) was quantified by a spectrophotometric method. The quantitative analysis of β-carotene production was achieved by measuring the spectra of β-carotene's characteristic λ_{max} peaks at 425, 450 and 478 nm. *E. coli P_{T5}-dxs P_{T5}-idi, E. coli P_{T5}-dxs* and the *E. coli* control strains were grown in 5 mL LB containing 25 μg/mL chloramphenicol at 37°C for 24 hr, and then harvested by centrifugation at 4,000 rpm for 10 min. The β-carotene pigment was extracted by resuspending cell pellet in 1mL of acetone with vortexing for 1 min and then rocking the sample for 1 h at room temperature. Following centrifugation at 4,000 rpm for 10 min, the absorption spectrum of the acetone layer containing β-carotene was measured at λ 450 nm using an Ultrospec 3000 spectrophotometer (Amersham Biosciences, Piscataway, NJ). The production of β-carotene in *E. coli P_{T5}-dxs* and *E. coli P_{T5}-dxs P_{T5}-idi* was approximately 2.8-fold

and 3.1-fold higher than that of the *E. coli* control strain, respectively (Figure 9). The production of β-carotene in *E. coli* P_{T5}-*dxs* P_{T5}-*idi* increased approximately 12% when compared to the parental strain *E. coli* P_{T5}-*dxs*.

EXAMPLE 9

Transduction of *E. coli* P_{T5}-*dxs* P_{T5}-*idi* with the P1 Lysate Mixture for Creation of a Better β-carotene Producer

E. coli P_{T5}-*dxs* P_{T5}-*idi* strain was transduced with P1 lysate mixture made on *E. coli kan*-P_{T5}-*lytB*, *kan*-P_{T5}-*dxr*, *kan*-P_{T5}-*ygbBP*, *kan*-P_{T5}-*ispA*, *kan*-P_{T5}-*ychB*, *kan*-P_{T5}-*gcpE*, and *kan*-P_{T5}-*ispB* strains (Figure 7). P1 transduction was performed as described in Example 7. Greater than 1000 transductants were produced. Among these transductants, 10 colonies that exhibited deeper yellow color than the parental strain *E. coli* P_{T5}-*dxs* P_{T5}-*idi* were chosen and the location of the insertion of *kan*-P_{T5} identified by PCR analysis as described in Example 7. PCR analysis showed that all ten colonies contained *kan*-P_{T5}-*ispB* as well as P_{T5}-*dxs* and P_{T5}-*idi*. *E. coli* P_{T5}-*dxs* P_{T5}-*idi kan*-P_{T5}-*ispB* strain exhibited deeper yellow color than the parental strain *E. coli* P_{T5}-*dxs* P_{T5}-*idi* on an LB plate containing both 25 μg/mL kanamycin and 25 μg/mL chloramphenicol after growing at 37°C for 2 days, suggesting higher yields of β-carotene production.

The kanamycin selectable marker from the chromosome of *E. coli* P_{T5}-*dxs* P_{T5}-*idi kan*-P_{T5}-*ispB* was eliminated as described in Example 6, yielding *E. coli* P_{T5}-*dxs* P_{T5}-*idi* P_{T5}-*ispB*. The elimination of the kanamycin selectable marker was confirmed by PCR analysis (Figure 8, lane 5 and 6). The specific primer pairs, T-kan (SEQ ID NO:35) and B-ispB (5'-ACCATAAACCCTAAGTTGCCTTT GTTCACAGTAAGGT AATCGGGG-3')(SEQ ID NO:39), T-T5 (SEQ ID NO:37) and B-ispB (SEQ ID NO:39) were used. Test primers were chosen to amplify regions located either in the kanamycin or the P_{T5} promoter and the beginning of the *ispB* gene (Figure 8). The PCR reaction was performed as described in Example 1. The PCR results (Figure 8, lane 5 and 6) indicated the elimination of the kanamycin selectable marker from *E. coli* chromosome and the presence of the P_{T5} promoter fragment in the front of chromosomal *ispB* gene based on the production of a band of the expected size (203 bp) which corresponds to the size of the P1 P_{T5} promoter sequence.

β-carotene production of *E. coli* P_{T5}-*dxs* P_{T5}-*idi* P_{T5}-*ispB*, *E. coli* P_{T5}-*dxs* P_{T5}-*idi*, *E. coli* P_{T5}-*dxs* and *E. coli* control strains was compared using the spectrophotometric method as described in Example 8 (Figure 9). The production of β-carotene in *E. coli* P_{T5}-*dxs* P_{T5}-*idi* P_{T5}-*ispB* was 3.4-fold higher than in the *E. coli* control strain (Figure 9). The production of β-carotene in *E. coli* P_{T5}-*dxs* P_{T5}-*idi* P_{T5}-*ispB* increased approximately 10% when compared to the parental strain *E. coli* P_{T5}-*dxs* P_{T5}-*idi*.

The present combinatorial P1 transduction method enabled, for the first time, isolation of the *ispB* gene capable of increasing the production of β-carotene under the control of the strong promoter. The isolation of *ispB* for increasing the production of β-carotene was an unexpected and non-obvious result because IspB, the enzyme octaprenyl diphosphate synthase, which supplied the precursor of the side chain of the isoprenoid quinones was expected to drain away the FPP precursor from the carotenoid biosynthetic pathway (Figure 1). The mechanism of how overexpression of the *ispB* gene, under the control of phage *T5* strong promoter, increased the β-carotene production is not clear yet. However, the result suggests that IspB may increase the flux of the carotenoid biosynthetic pathway. Combinatorial transduction is a powerful tool for the identification of new genes in biosynthetic pathway optimization.